nature

20 August 1981

Saving antibiotics from themselves

For how much longer will antibiotics remain at the disposal of physicians? This is the question that will be raised in the public mind by the gloomy statement put out earlier this month from Boston by a group of participants in the conference on the Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids held in Santo Domingo last January. The statement, in which Professor Stuart B. Levy of Tufts University has played a prominent part, points to the spread of resistance to antibiotics among bacteria in recent years, says that the "worldwide public health problem" is "due in large part to the indiscriminate use of antibiotics" and goes on to urge that "national and international committees" should be set up so as to issue "directives for prudent antibiotic use". Dr Levy's group is anxious that attention should widely be paid to its clarion call.

Nobody dissents. The problem of antibiotic resistance, although something of a surprise because thirty years ago nobody anticipated the versatility of bacterial plasmids, is, nevertheless, not new. Good sense argues in the same direction as Dr Levy's group: plainly everybody — not merely physicians but public health officials and farmers — should be more discriminating in the use which they make of a class of compounds, for the time being invaluable, whose value is being eroded as the years go by. Less vitally (in the human sense), the same process has been happening with insecticides, the other principal biochemical legacy of the 1940s. Organisms of all kinds, not only bacteria but insects as well, constantly surprise us by their biochemical versatility. (So, too, do higher organisms such as human beings, which have demonstrated a remarkable capacity for survival in circumstances in which self-adulteration with materials such as alcohol is common, although there is no reason to suppose that selection has played any significant part in this process.) How do we learn to live with such surprises?

The question is every bit as important as the statement earlier this month makes out. It is more than merely wasteful, but tragic as well, that the benefits of technological innovations (antibiotics, insecticides) should irreversibly be misused. Superficially, these sombre prospects are more galling than the waste of the potential benefits of mechanical technology — air transport (hamstrung by protective regulation), telecommunications (often hamstrung by protective legislation) and computer technology (hamstrung by collective inertia). For there, the argument goes, the promise remains, waiting merely on some kind of reform of people's resolve to make sensible use of the tools that science and technology have provided for them. But is it not also tragic that in present circumstances the productive wealth of the industrialized nations should be constrained, with all the human consequences that are entailed — and measured by the growing proportions of the unemployed — and that the gap between the prosperity of the rich and the poor should, through mutual impoverishment, be sharpened? Society, in other words, is as much at risk from the under-use as from the over-use of the technology at its disposal.

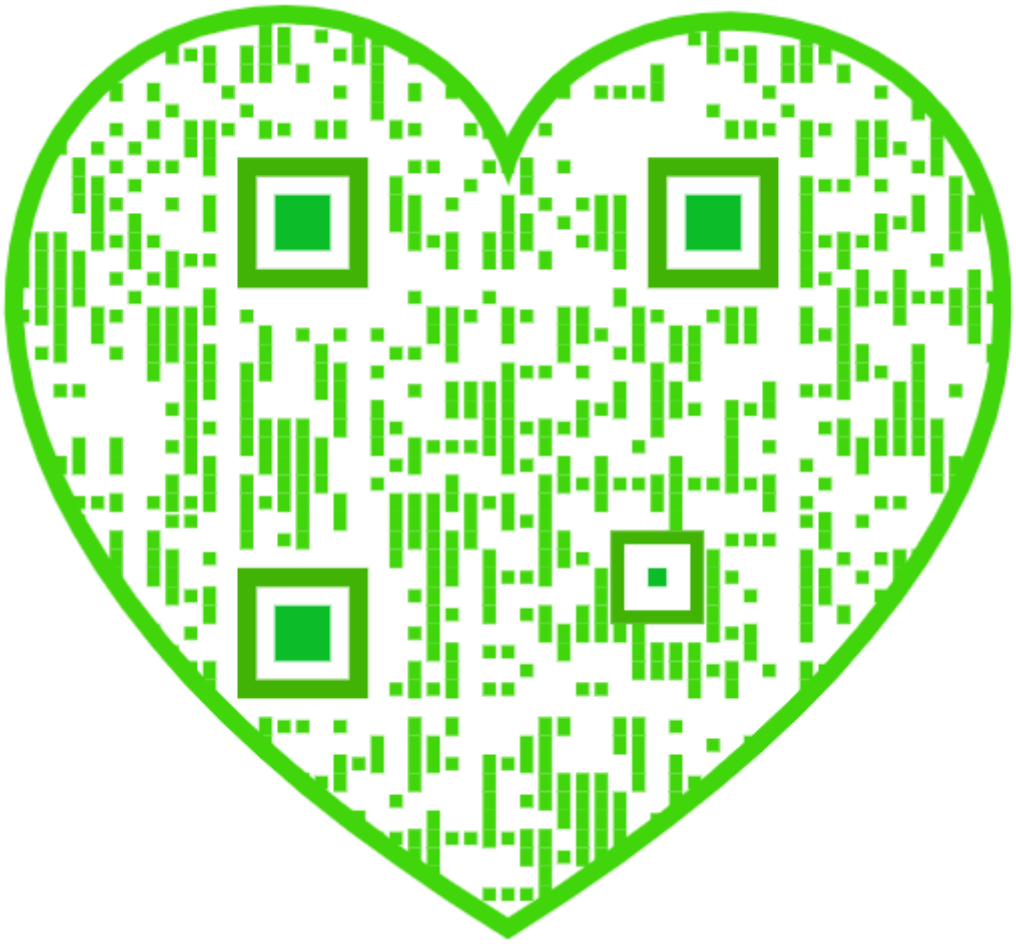
How, against this background, should enlightened policies on antibiotics be designed? Dr Stuart Levy's group, recognizing that policies of some kind need to be devised, has been plainly at a loss to know how these might come into being. Why else would it have suggested a hierarchy of national and international committees? This strategem, the most obvious, is well calculated to amplify its own eloquent statement of foreboding about the consequences of the misuse of antibiotics, but unlikely to make much headway with dealing with the causes of misuse. These, in turn, are deeply embedded in the ways of societies blessed with the chance of using antibiotics — the freedom of physicians to prescribe, the freedom of other professional people (farmers, for example) to use chemical products in ways that they think proper and profitable and the general restraint on democratic government from excessive interference with what people do (or do not do). Dr Levy's group, for all the best reasons, finds itself up against the kind of social ethos that made the discovery of antibiotics possible. That, it may be thought, is the irony of their (and our) dilemma. How is it to be resolved?

The dismal science would suggest a simple remedy, not entirely to be scorned. If some uses of a potential social benefit are judged democratically to be prejudicial for society at large, why not put a tax on them? The snag is that the demand for antibiotics is greatest in those societies in which the prices of supposedly life-saving drugs have the least negative effect on the demand for them; elsewhere, where the "need" is perhaps greatest, the "demand" (muted by the capacity to pay or even by ignorance) is least. An alternative solution may be found in the common ethics of the medical profession, which increasingly requires that physicians should balance the benefit that a single patient will receive from a certain treatment against the more general benefit if the resource in question is dispensed in some other way — the allocation of a kidney machine to one person rather than another, for example. In the use of antibiotics, the counterpoise of individual benefit is, however, non-use for the greater good. Logically, the problem is the inverse of that in which physicians subject their patients to the small risk of damage from the use of a vaccine (coupled with the benefit of immunity against the corresponding infection) for the sake of the wider benefit. The trouble is that non-use is certain to be regarded, in democratic societies, as deliberate and thus reprehensible neglect. How, in the last resort — as it may be — does a physician explain that the administration of an antibiotic likely to be beneficial must be withheld for fear that resistance will be encouraged?

Physicians properly concerned about problems of antibiotic resistance must manage more adequately to come to terms with the societies in which they work. It would, of course, in many other fields than this, be splendidly simple if individuals would put their expectation of personal benefit second to the communal benefit, even in matters of life or death. In the short run, however, that is exceedingly unlikely. And there is no immediate prospect that there will quickly arise such a persuasive analyst of the machinery of society that some compromise will be found between the doctrines of the two greatest analysts so far — Adam Smith (who can be bowdlerized as holding that the communal good is the sum of individuals' estimates of their individual benefit) and Karl Marx (whose essential argument, grossly oversimplified, is that the greater good will benefit most but not all individuals). So what should physicians do about antibiotics? The hierarchy of committees now proposed, potentially useful, should be given positive rather than negative terms of reference. The objective should be somehow to suggest things to do rather than not to do — handlists for physicians to help them to make empirical decisions, programmes for the more effective control of infection in hospitals, plans for helping forward the development of new kinds of antibiotics (for research people are even more adaptable than bacteria) and, above all, studies of how a proper balance between use and non-use is to be struck in modern democratic society. Nobody will pretend that these tasks are simple. In the long run, however, nobody will listen seriously to Dr Levy's group unless it tackles these broader subjects.

0028-0836/81/340661-01$01.00

© 1981 Macmillan Journals Ltd



London bites private bullet publicly

University cuts spending across the board

The University of London last week let its constituent colleges know how much they will have to spend next year, and how many students they will be permitted to educate three years from now, in 1983–84. The target reductions of the student population are in line with the gloomier forecasts; most colleges are disappointed with their financial allocation for the academic year immediately ahead.

Some, however, are relieved. Earlier this year, Chelsea College, one of the science-based smaller colleges in the university, was singled out by the second interim report of the Committee on Academic Organization commissioned by the university for oblivion or dismemberment; Chelsea is therefore understandably relieved that the now threatened reduction of its student numbers three years hence is a mere 14 per cent.

Other colleges, previously thought to be set fair, have been downcast. Queen Elizabeth College, for example, applauded in the most recent interium report of the Committee on Academic Organization (under Sir Peter Swinnerton-Dyer) has been asked to plan for a student body reduced by a third three years from now. The reduction is even more drastic than that decred by the University Grants Committee for the University of Aston in Birmingham.

Under the university's plan for the cutting of the cake for the coming academic year, the larger colleges do best. University College London, singled out for its high unit costs, by the Swinnerton-Dyer Committee, will have to stomach only a five per cent cut in student numbers in the next five years.

In its letter to schools of the university last week, the Court said that it had decided to take full account of the recently decreed increase of fees for part-time students, but had made a special but interim provision for Birkbeck College (most of whose students are part-time). On overseas students (or the lack thereof) the Court said that it could not undertake to help those branches of the university hard-hit by recent changes.

The Court's letter contains an element of second guessing in the statement that it has "devised arithmetical values for the various UGC verbal indications about Arts and Sciences". The upshot of the Court's calculation seems to be that, after allowing for the buoyant future spelled out for Imperial College, the population of science students in the rest of the university will decrease by 11.3 per cent in the next three years, compared with the reduction of 5.7 per cent expected in other academic fields.

The university court seems, nevertheless, to have been perplexed by some of the consequences of its arithmetic. Given the substantial reduction of student numbers in science foreseen, and the grants committee's recommendation (to the university as a whole) that special attention should be given to veterinary science, mathematics and engineering, the Court records that it enquired whether a reduction was required in biology.

Apparently it "emerged" that the committee had not intended that the subject group of "biological sciences" should "suffer such a large reduction" but that the committee was in favour of a reduction of the numbers of students following degree courses combining two science subjects, "a category not specifically mentioned in the UGC letters".

On balance, the Court has planned for a real reduction of 5 per cent in its resources in the coming academic year. Given previous commitments to medical and other professional schools in the university, the Court says that the generality of its dependants will have to make do with 7.5 per cent less in the coming academic year.

Soviet Union urges space demilitarization

The Soviet Union has proposed a new treaty banning the deployment of weapons in outer space. The terms of the treaty, set out in a letter from the Soviet Foreign Minister, Andrei Gromyko, to the Secretary General of the United Nations, Dr Kurt Waldheim, differs from the 1967 treaty on outer space in two important respects. It covers all weapons, and is envisaged as an instrument of the United Nations, to be available for signature by "all states" with the instruments of ratification deposited with the United Nations Secretary General and not, as in 1967 with individual sponsoring governments. Accordingly, in the opening clauses of the draft, the text, refers to "member states" and switches to "signatories" only in details dealing with the technicalities of ratification, amendment and possible withdrawal. According to TASS, the essential articles of the draft read as follows:

> 1.1 The member states undertake not to put into orbit around the Earth objects bearing weapons of any kind, not to install such weapons on celestial bodies, and not to deploy any weapons in outer space in any other way including on piloted spacecraft of multiple use both of the existing type and also of other types which may be developed in the future by member states.
>
> 1.2 Each member state . . . undertakes not to help, encourage and incite any state, group of states or international organization to engage in activities running counter to the provisions of paragraph 1 of the present article.
>
> 2. The member states shall use spacecraft in strict accordance with international law including the UN Charter in the interests of maintaining international peace and security, and for the development of international cooperation and mutual understanding.
>
> 3. Each member state shall be bound not to destroy, damage or disturb the normal functioning or alter the flight trajectory of space vehicles of other signatories.
>
> 4. For the purpose of ensuring confidence in the observation of the provisions of the present treaty, each member state shall use the national technical monitoring facilities at its disposal in such a way as to comply with the generally accepted principles of international law.

The document is, of course, only a draft, and with the United States toying with orbital laser weapons seems unlikely to achieve anything concrete in the near future. The US shuttle, however, and its possible military implications have been causing considerable concern to Soviet military experts. The past few months have seen numerous comparisons in the Soviet media between the "militarist" capabilities of the shuttle and the essentially peaceable nature of the Soviet Salyut orbital stations. Such concern seems, at least in part, to lie behind Mr Gromyko's letter to Dr Waldheim, which notes that the risk of the militarization of outer space has "recently increased".

If the treaty should materialize, several ambiguities in Mr Gromyko's draft would doubtless have to be resolved. Thus article 3, as it stands, seems to leave open two justifications for damaging or disturbing a spacecraft in flight — that it was not (or was erroneously thought not to be) the property of a member state, or that it was allegedly launched in breach of article 1.1. Again, the draft contains no clear definition of a 'weapon". In its strictest sense, article 1.1 would preclude even the carrying of a hand gun or life-preserver by, say, senior officers of a large space station. Such a station, too, might well wish to deploy small rockets with explosive warheads to destroy artificial debris or meteors on a dangerous approach course. Would such rockets be classed as "weapons" if they were "accidentally" launched against the spacecraft of another signatory? Moreover, just as in 1967, the treaty makes no provision for inspection, simply calling (article 4.3) for consultations and exchange of information "where necessary". Without such provisions, any such treaty would be little more than an ethical gesture.

Vera Rich

0028-0836/81/340662-02$01.00

©1981 Macmillan Journals Ltd

UK biotechnology

Fast start

Britain's "official" biotechnology company, Celltech Limited, seems not yet to have decided how to respond to Stanford University's advertisement of terms for non-exclusive licences to exploit the Boyer–Cohen patent. Part of the difficulty is that it is not yet clear how far companies outside the United States, part of whose business will come from the licensing of their own patent rights to other non-American companies, will be affected by the patent. But there is also an element of waiting and seeing. So far there have been no direct talks with Stanford.

In any case, the company is probably too busy. Since Celltech moved into its new headquarters in Slough (20 miles west of London) just after Christmas, the staff has grown more quickly than originally foreseen, to just about sixty. The company seems pleased with the eagerness of scientists to join up, mostly from industrial companies but with a substantial proportion of people from the universities.

One of the newest and most significant recruits is, however, a sales invoice clerk. Although the company's prospectus has made no bones about the time that will be needed to put biotechnology on the commercial map, it does plan to sell what it can as soon as it can. One of the first products will be an anti-interferon monoclonal antibody, principally for use by the manufacturers of interferon. A laboratory for the preparation of this and other monoclonal antibodies is being assembled. But the company is anxious that this part of its business should not be thought of as more than a kind of sideline.

Understandably, the company is reluctant to say where its chief efforts will go, but gives the impression that it will not be investing a major effort in the production of interferon of any kind. At some stage, though, it seems clear that Celltech will be in the industrial enzyme business. For the time being, the practical aim seems to be to establish itself well in the basic techniques of genetic manipulation, among which nucleic acid chemistry and the techniques of gene expression are conspicuous.

The nucleic acid chemists have built themselves an automatic nucleotide assembly machine rather than buying one off the peg, chiefly for the flexibility that results. They claim that their machine will add nucleotides (up to a score or so) at the rate of one every half an hour, which means the production of one full-length oligonucleotide during each overnight run.

Celltech's fast start means that its initial capital of £12 million will be used up more quickly than originally foreseen. The plan now is to recruit up to 150 people by September 1983, half of them scientists.

The shareholders in the company (among whom the National Enterprise Board is the largest, with 44 per cent) have agreed to the accelerated investment plan, and will take a close look at its successor in just over two years time. The officers of the company say that they are heartened by the long-term view taken by the shareholders, which also include the Prudential Assurance Company, the Midland Bank, British and Commonwealth Shipping and Technical Development Capital. This, the guess is, should give Celltech a good chance when, in two or three years, there is likely to be a shakeout in the biotechnology field.

The company has no fixed ideas about the way in which it will sell its products, but plans not to enter the marketing of manufactured products nor manufacturing as such (which does not exclude the setting up of manufacturing companies in partnership with others). Enquiries from industry have been flowing in, some of which may lead to research contracts.

On the relationship between Celltech and the Medical Research Council, on whose developments the company has first refusal, Celltech says that developments not of direct interest may be licensed to others, in which respect the company will act as a kind of broker. It expects, however, to make direct use of most of the ideas suggested by the Medical Research Council, which, nevertheless, retains the right to bypass Celltech if too much time passes.

US graduate salaries

Engineers ahead

Washington

Statistics for 1980 recently released by the US Department of Education show a significant increase in the number of bachelors and first professional degrees in science and engineering awarded by US colleges and universities for the first time since the early 1970s.

The main reason for the increase is a growth of 11.5 per cent in the number of engineering graduates, continuing a trend that started in 1976. Since then, the rise has been balanced by a drop of about 4 per cent a year in the number of graduates in the social sciences and psychology — there was a drop of only 1.2 per cent in 1980.

The continued growth in engineering graduates reflects the relatively high salaries that even first-degree graduates are being offered by industry. Figures published by the College Placement Council two weeks ago showed that 1981 graduates are entering jobs with starting salaries between 10 and 14 per cent higher than a year ago.

The average starting salary for petroleum engineers with a bachelor's level degree was $26,652, and for chemical engineers $24,360. This compares with starting salaries for economists with similar qualifications of $16,440, and for other social scientists of $13,992.

The high salaries being offered by industry also seem to be discouraging engineering graduates from staying on to complete higher degrees, leading to difficulties in higher education institutions of attracting qualified teaching and research staff. In comparison with the 11.5 per cent increase in bachelors' and first degrees awarded in engineering, the number of masters degrees in the same subjects rose by only 4.3 per cent, and the number of doctors' degrees remained virtually constant.

Concern about the implications of these figures has prompted Congress to oppose President Reagan's plans to eliminate the National Science Foundation's support for graduate fellowships. Committees in both the House and the Senate have agreed to restore funds for the fellowship programme, which makes about 500 new awards a year, although the precise level of funding for the foundation's overall education efforts has still to be negotiated between the two legislative bodies.

Another problem now emerging is a shortage of qualified science teachers. States have experienced a rapid decline in the number of science teachers turned out by their colleges of education; in physics, for example, the supply of new secondary school teachers has virtually dried up at a time when warnings are being voiced in Washington about the dangers of a growing "scientific illiteracy" among the general population.

According to Dr James Rutherford, who has been appointed director of a new science education programme within the American Association for the Advancement of Science, many schools are reporting that their science teachers are being approached with tempting job offers from private companies — often mathematics and physics teachers are offered jobs in computer sciences at salaries considerably higher than they receive for teaching.

David Dickson

Degrees awarded in 1980 (figures from National Centre for Education Statistics)

	Bachelors' and first professional degrees		*Masters' degrees*		*Doctors' degrees*	
Total	1,009,879	(+0.9%)	299,492	(−0.9%)	32,758	(—)
Science and engineering	292,271	(+1.2%)	54,463	(—)	16,495	(+0.8%)
Science and mathematics	117,906	(−1.1%)	22,023	(−1.2%)	8,559	(+1.0%)
Social sciences and psychology	114,469	(−1.2%)	15,552	(−2.7%)	5,416	(+0.8%)
Engineering	59,903	(+11.5%)	16,888	(+4.3%)	2,520	(+0.1%)

Values in parentheses are per cent changes from 1979 figures; a dash indicates no change.

0028-0836/81/340663-02$01.00

©1981 Macmillan Journals Ltd

Indian health care

Keeping it simple

Lucknow

If India's ambitious target of "health for all" by the year 2000 is to be approached, then the present curative hospital-based, urban-centred health service may have to give way to simpler low-cost alternatives, especially in rural areas.

Some advances have been made in developing "appropriate technology" for health care. Low-cost audiovisual aids and teaching methods for various levels of health worker have been developed under the aegis of the Indian Council of Medical Research. The Central Scientific Instruments Organization has developed a cycle-ambulance with a stretcher for the patient and a seat for the attendant; a clinical thermometer with temperature ranges indicated as coloured bands for easy use by illiterate health volunteers; a simple apparatus for vasectomy, now ready for human trial; a fetal stethoscope which can detect a fetal heart beat at week 12 of pregnancy; a miniature cardioscope and a colorimeter. The Vector Control Centre, Pondicherry, has demonstrated that certain fishes such as *Aplocheilus blochii* and *Gambusia affinis*, can be used as effective controls of mosquitoes, on whose larvae they feed.

For maternal and child health care by traditional birth attendants, called *dai*, a simple *dai* kit has been developed at the Postgraduate Institute of Medical Education and Research, Chandigarh. The kit should reduce the risks of tetanus, by achieving better asepsis, and helps the birth attendants to spot mothers and infants at risk — again using colour-coding principles on measuring sticks and weighing scales.

A much-needed boost to rural health care should come with the recent decision of the Indian government to set up a national committee on appropriate technology for health care. The Indian Council of Medical Research is to provide technical support for the national health programme — and at a national workshop held recently by the council the value of community health volunteers and multipurpose health workers was underlined. **Zaka Imam**

Blood products

Surplus sale

Products extracted from donations to the British National Blood Transfusion Service, and for which the National Health Service has no use, may in future be sold to pharmaceutical companies, the government has decided. Although some companies have occasionally bought products from the transfusion service in small quantities, usually for research purposes, there have been no formal arrangements for buying in bulk. Thus British companies have had to buy from abroad.

The decision to sell off surplus blood products has emerged from a lengthy wrangle over the use of products that are now discarded, expected to be exacerbated as the transfusion service puts into operation a plan to make more blood plasma available for fractionation. There has also been a long discussion about the revamping and expansion of the Blood Products Laboratory, which carries out all fractionation of donated blood. The laboratory suffers from some of the problems of the former Lister Institute of Preventive Medicine, whose site it shares. The institute was closed in 1978 after it began to lose money and could not afford to modernize its facilities.

An earlier government suggestion that the Blood Products Laboratory be handed over to commercial management fell flat after many organizations, including blood donor groups, trades unions and Labour Members of Parliament, objected to the rise of blood donated by volunteers for overtly commercial purposes. As a compromise, the government has decided to keep the laboratory firmly under the wing of the Department of Health and to sell products surplus to the National Health Service's requirements. Any profits will be ploughed back into the health service, although it is not yet clear whether they will be spent on updating the laboratory.

The success of the scheme will depend on whether the laboratory can supply companies with products they need. Companies are negotiating with the health department, which has said only that outdated red cells and plasma fractions containing nonspecific antibodies will be available. Products for which the health service's demand exceeds supply, however, will not be for sale. They include factor 8, the blood clotting agent used for treating haemophiliacs, and albumin.

Future sales will, however, depend on the plan to make more blood plasma available for fractionation, by substituting concentrates of red cells for whole blood in transfusions and on the laboratory's ability to take on the extra work. The aim is that by the end of 1982, the laboratory will have enough plasma to produce two-fifths, compared with one-fifth at present, of the national demand for factor 8, in which it is hoped that Britain will ultimately be self-sufficient. That will mean that other products, such as factor 9, immunoglobulin and albumin, will be produced in excess and will be available for sale.

One company now discussing its needs with the Department of Health is Beechams Pharmaceuticals Limited, which would like to buy plasminogen for clinical trials in the treatment of deep-vein thrombosis. Whether the laboratory will be able and willing to supply such specific products, or will simply make certain plasma fractions available, remains to be seen. **Judy Redfearn**

Paying court

Forensic scientists in Britain are rather touchy at the moment, following the suspension of one of their number for allegedly failing to reveal crucial ambiguities in evidence on blood samples, and now the "demotion" of another for writing a research paper which is said to reveal detachment.

This second scientist resigned rather than accept a teaching post — which he regarded as demotion — and then went to an industrial tribunal to claim what he saw to be effectively unfair dismissal. However, the tribunal rejected the scientist's claim, and he is to appeal to the Employment Appeal Tribunal in London on 3 September.

His paper, which appeared in the official journal of the British Academy of Forensic Science — *Medicine, Science and the Law* — in 1977 (vol. 17 no. 1, p.37), deals with the quantitative toxicology of corpses — addressing the issue of how to estimate a drug or poison dose. The paper uses a simple law of linear diffusion to relate the drug content of various organs (taken from the data sheets of real cases) to dose and time of death as estimated by the police and witnesses. The data are clearly very scattered and the fit loose, and the author concludes "A pessimistic view is that it is impossible to remove the last vestiges of alchemy from [forensic] toxicology because of the nature of the material for analysis. It may be, apart from the fact of death, that one cannot be absolutely certain of anything and there are always contrary authorities to be found".

This conclusion may not have recommended itself to the Forensic Science Service, which as part of the Home Office runs forensic science in the United Kingdom. Other forensic scientists may equally have been unsettled.

Robert Walgate

US uranium

Falling demand

Washington

A report from the US Department of Energy (DoE) has revealed just how depressed the domestic market for uranium has become. This was already evident from the dramatic price reduction in 1980, when uranium oxide fell from $43 to less than $30 a pound. Now DoE has shown that, for the first time, the amount of new orders for future supplies placed by US utilities was more than outweighted by cancellations of existing commitments.

The DoE's report, *A Survey of United States Uranium Marketing Activity*, published last week, is based on information gathered from 60 utilities with nuclear reactor projects, 30 present or potential uranium users and five reactor manufacturers. It says that domestic

0028-0836/81/340664-02$01.00

©1981 Macmillan Journals Ltd

buyers contracted with US primary producers for an extra 12,500 tons of U_3O_8 in 1980, but that this was offset by a reduction of 25,600 tons for commitments made under previous contracts.

The survey showed that, despite the current depressed prices, most utilities and producers expect the price of uranium to rise sharply over the next decade; future contracts which guarantee a floor price for any 1 year quote an average of $75 a pound by 1990.

The drop in uranium demand is blamed on the extended licensing period faced by companies building new reactors, a reflection of increasingly stringent safety requirements and on the substantial reduction in estimates for future demand for electricity as the United States becomes more conservation-conscious, and energy costs continue to rise.

Some business analysts feel that the price of uranium may have bottomed out, at least temporarily. According to calculations by the Californian consulting company Nuclear Exchange Corporation (Nuexco), uranium prices remained virtually constant at $24 a pound between April and June, and actually increased slightly last month.

However, the slump in prices — and demand — is clearly reflected in difficulties being faced by the uranium mining industry. Some estimate that as many as 4,000 workers, 25 per cent of the workforce, have been laid off.

This decline, too, is reflected in DoE statistics. In another report, soon to be published, the department notes that the uranium mining industry is entering a period of major retrenchment. Having initially planned to spend $316.5 million on new mining facilities in 1981, for example, the industry now estimates that total spending will be only about $172 million. Projections for capital spending on mill facilities have dropped even more dramatically, from $241.9 million to $99.3 million.

According to DoE officials, several processes are keeping the price of uranium oxide down. One is that many utilities, faced with high interest rates, are selling stockpiled uranium back to producers. Another is that, even though uranium mining has been cut back, production figures are still going up as companies work through stockpiles of ore. Nuexco had predicted a drop of 2.5 million pounds between 1979 and 1980 to about 34.5 million pounds — still about double the amount consumed by the nuclear industry.

In fact, however, production increased by 13 per cent to 42.5 million pounds. Some wastes which have already passed through the mill are being rerun, a way of obtaining more uranium oxide without having to pay for any extra mining. Some industry analysts feel that this will continue, so that even if demand picks up it will not be reflected by an immediate boost for the mining industry. **David Dickson**

German cancer research

Politics ousts science

Brussels

The scandal at the German Cancer Research Institute has grown into a national issue. The scientific director has resigned, the administrative director appears to have been removed, the oppositions parties (CDU/CSU) have attacked the government and the institute itself is now involved in internecine strife.

The national research centre in Heidelberg is a union of eight institutions centrally administered by the Bundes Ministerium für Forschung und Technologie (BMFT), with a 1981 budget of DM 79 million ($1,817 million). The latest scandal began with the appointment of Dr Hans Neurath, born in Vienna, but previously professor of biochemistry at the University of Washington, as the centre's scientific director. The post had been vacant since 1976 and the chief administrator, Bodo Spiekermann, was relieved to have found someone with the qualifications and drive to attempt the much needed reorganization.

The other directors of the centre were not consulted about the appointment and soon raised a storm about the professor's demand that the institute pay for a villa in addition to his salary of DM 167,000 a year. The attack soon broadened to Neurath's qualifications and it turned out that he had in fact previously only been scientific advisor to the University of Seattle.

The centre's director for toxicology and chemotherapy, Professor Dietrich Schmaehl, declared that Neurath could not possibly be sufficiently knowledgeable about cancer research to take over leadership of the centre. Nothing daunted, Neurath set about appointing outside investigators to assess the value of the research projects in hand with the intention of directing resources to the most promising areas. The multidisciplinary, clinically-orientated approach was rejected, and Neurath, with the support of the ministry's director, Wolfgang Spinke, proposed to cut back on activities such as nuclear medicine and chemotherapy.

The zeal with which Neurath set about reforming the institute inflamed the already poisoned atmosphere, the battle amongst the divided researchers spilled over into the press and by May this year it had become a political issue, with the opposition, CDU/CSU, siding with Neurath's opponents and the government defending its appointee who by then had been in charge for a year.

The opposition claimed that the administration of the centre was chaotic and scandalous. By July CDU/CSU was describing the working climate as completely destroyed by internal power struggles complete with the "terrorization of administrative personnel".

On 24 July, Neurath announced that he had resigned but would explain the reasons for his decision only in September. BMFT also announced that Spiekermann was being withdrawn, and the conflict became an out and out battle between the ministry and the research centre. The ministry now plans to examine the scientific activities of the centre since 1973.

The latest chapter in the saga is the government's reply to the CDU/CSU's accusations in July. It reaffirms BMFT's desire to strengthen pure research rather than to concentrate on already well-explored areas. In addition, the government says it has long been convinced that the quality of the centre's research could be improved and that the routine work carried out runs counter to the aims of the centre. "Internal egoism", it says, has been responsible for the present situation; the attack against Neurath is merely an attempt to divert discussions from the real issues at stake.

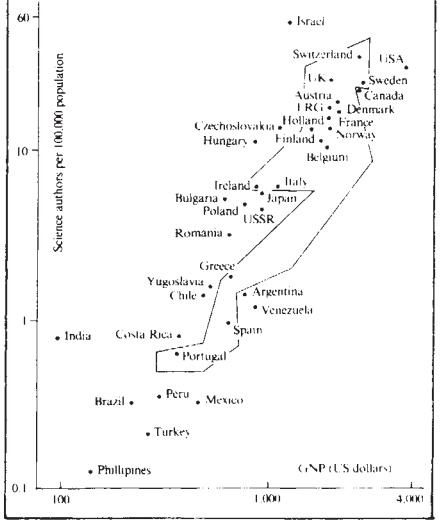
Jasper Becker

Productivity by numbers

India, Eastern Europe and Israel stand out as being unusually scientific in the accompanying graph prepared by sociologist Dr Pedro Gonzalez-Blasco of the University Alcalá de Henares, Madrid.

The plot shows the number of scientists per hundred thousand population against gross national product (GNP) per capita in 1967. The box encloses Western European countries. The main trend is for the proportion of scientists to rise at about the third power of GNP per capita — so that scientists are added more rapidly per unit of additional GNP the higher the GNP. However, this may be read in two ways. Are the countries above the trend more effective at creating scientists? Or are their scientists less effective at creating wealth?

Dr Gonzalez-Blasco's plot forms part of the background to a seminar on the state of physics in certain less developed European countries to be held in Istanbul on 4–5 September under the auspices of the European Physical Society.



0028-0836/81/340665-01$01.00

© 1981 Macmillan Journals Ltd

CORRESPONDENCE

Badger vaccination

SIR — It was noble of Mr Rees (*Nature* 9 July, p.104) to come to the defence of the Zuckerman report; however, my criticisms were of the report rather than the Ministry.

Mr Rees misreads my plea that some attention should be paid to vaccination of badgers, as advocacy of vaccination as *the* answer.

It is well known that BCG is one of the less effective vaccines in man, both in the percentage of the population protected, and in the type of protection afforded (infection is not prevented, it is merely limited). Reports have shown that protection can vary from 78% of the vaccinated (human) population at best to negligible responses at worst. However, protection of only some of the population is epidemiologically significant, and whilst protection may not be absolute it is of proven value in reducing the transmission to further individuals — the significant factor here. The epidemiological basis in controlling a disease is to reduce the number of individuals infected by each diseased animal in its lifetime to less than one. Therefore any means of reducing the number of susceptible individuals contacted by each infected badger will be of significance. (Gassing undoubtedly comes into this category).

Though attempts to vaccinate foxes orally (against rabies) have shown what problems there can be, badgers are well known among zoologists for the ease with which they will take quite considerable quantities of easily prepared oral baits. This has been used to advantage in a variety of ecological studies.

The degree of infection in some areas was also cited as a reason against vaccination, and whilst gassing may be the most effective approach in severely affected areas, it is nevertheless still the case that vaccination alone, if feasible, would have some positive effect. What does not appear to have been considered was the value of vaccination in less infected areas, or even ring-vaccination around those areas to be gassed (migration into cleared areas being a problem).

At this stage all conjecture is limited by the lack of knowledge. However, I must still regret that the report did not more positively specify research aims. I still believe vaccination to be one of these.

M.J. CHAPMAN

Easingwold,
York, UK

White or brown?

SIR — Concluding his review of two important books of contemporary scientific and medical interest on dietary fibre and health[1], John Yudkin says "Let us by all means eat brown or wholemeal bread if we like it. But let us not delude ourselves that it will make us healthier or prolong our lives". This statement is not justified by the data available on the efficacy of dietary fibre in lowering disease risk. There may be useful effects of certain dietary fibres in preventing or ameliorating intestinal diseases in general, but this aspect is for others more expert to discuss[2,3]. We now comment on the scientific aspects of fibre use in relation to colon cancer.

Countries with a high incidence of atherosclerosis and coronary heart disease usually have a high incidence of colon cancer, and vice versa[4-6]. For example, Great Britain has a high incidence of both kinds of disease, the Japanese a low incidence. Finland is an exceptional country, exhibiting a high incidence, in fact among the highest in the world, of atherosclerosis and myocardial infarction; however, interestingly, their colon cancer incidence is only slightly higher than that of Japan, but much lower than the other Scandinavian countries[7,8]. Independent groups in the United States and Europe have investigated the underlying mechanism[9,10]. Detailed inquiry showed that the Finnish people consumed a diet not too different with regard to the amounts of most components, especially lipids and proteins, from those of people in New York or Copenhagen, so, by current concepts their diet and nutrition account for their high incidence of heart disease. However, the Finnish people also consumed rather large amounts of whole grain bread and other sources of cereal fibre, resulting in a stool bulk two to three times larger than those of the populations in New York or Copenhagen. At the level of metabolic epidemiology[9,10] specifically in relation to secondary bile acids which were demonstrated to be promoters in the carcinogenic process[11], the people in Finland and New York showed the same total bile acid levels, as would be expected with similar lipid intakes. However, and importantly, because of the increased stool bulk in Finland, the *concentrations* of bile acids were much lower — similar to the concentration found in Japan. The Japanese level was that low, not because of high fibre intake, but because of low fat intake[12]. Thus, the epidemiology was explained by metabolic biochemical approaches, and it was demonstrated through studies in man that cereal fibre accounted for a decreased risk of colon cancer in Finland. Furthermore, parallel studies showed that cereal fibres, particularly bran, could decrease the incidence of colon cancer in animal models[13,14]. The few instances where this was not so involved feeding the fibre during the period when carcinogen was administered, but not during the important promotion phase[15], or when too high a level of carcinogen was administered, thus overwhelming the promotion phase where cereal fibre exerts its major protective role[16].

To be sure, more research is necessary on the effect of diverse fibres, and in fact on the effect of other micronutrients. Unlike Yudkin, we believe that the present data base is adequate to conclude that brown or wholemeal bread with adequate fibre and other whole grain cereals, such as bran-containing breakfast cereals, are not only tasty but, indeed, have every chance of reducing the risk of colon cancer, a major type of cancer in the Western world. It is so easy to implement, and in fact would simply be a return to the situation prevailing before extensive milling of cereals removed much of the valuable fibre. Thus, adverse health effects owing to possible mineral imbalances are unlikely. The consequences in lowering the risk of colon cancer, possibly other intestinal disease and even early diabetes[17,18] would be appreciable if not "dramatic". Please, Dr Yudkin, be objective.

J.H. WEISBURGER
B.S. REDDY

American Health Foundation,
Naylor Dana Institute
for Disease Prevention,
Valhalla, New York, USA

1. Yudkin, J. *Nature* **291**, 173–174 (1981).
2. Burkitt, D.P. & Trowell, H.C. (eds) *Refined Carbohydrate Foods and Diseases* (Academic, London, 1975).
3. Burkitt, D.P. *Am. J. clin. Nutr.* **34**, 428–431 (1981).
4. Blackburn, H. *Prev. Med.* **8**, 612–678 (1979).
5. Winawer, S.J., Schottenfeld, D. & Sherlock, P. (eds) *Colorectal Cancer: Prevention, Epidemiology, and Screening* (Raven, New York, 1980).
6. Turpeinen, O. *Circulation* **59**, 1–7 (1979).
7. Waterhouse, J., Muir, C., Correa, P. & Powell, J. (eds) *Cancer Incidence in Five Continents* Vol. III (Int. Union Against Cancer, Berlin, 1976).
8. Teppo, L. & Saxen, E. *Israel J. med. Sci.* **15**, 322–328 (1979).
9. Reddy, B.S., Hedges, A.R., Laakso, K. & Wynder, E.L. *Cancer* **42**, 2832–2838 (1978).
10. IARC Intestinal Microecology Group *Lancet* ii, 207–211 (1977).
11. Reddy, B.S., Cohen, L.A., McCoy, G.D., Hill, P. & Weisburger, J.H. *Adv. Cancer Res.* **32**, 237–245 (1980).
12. Hirayama, T. *Nutr. Cancer* **1**, 67–81 (1979).
13. Reddy, B.S., Mori, H. & Nicolais, M. *J. natn. Cancer Inst.* **66**, 553–557 (1981).
14. Nigro, N.D., Bull, A.W., Klopffer, B.A., Pak, M.S. & Campbell, R.L. *J. natn. Cancer Inst.* **62**, 1097–1102 (1979).
15. Bauer, H.G., Asp, N., Oste, R., Dahlquist, A. & Fredlund, P. *Cancer Res.* **39**, 3752–3756 (1979).
16. Cruse, J.P., Lewin, M.R. & Clark, C.G. *Lancet* ii, 1278–1280 (1978).
17. *Lancet* ii, 423–424 (1981).
18. Manhire, A., Henry, C.L., Hartog, M. & Heaton, K.W. *Lancet* i, 1157 (1981).

Yudkin's answer

SIR — My article, to which Weisburger and Reddy refer, was a review of two recent publications, both entitled *Medical Aspects of Dietary Fibre.* One of these was a report of the Royal College of Physicians, which had this to say about fibre and colon cancer:

> "There are reasonable grounds for the statement that, in genetically susceptible persons, large bowel cancer could be favoured by a fibre-depleted diet, but other explanations for the commonness of this cancer in Westernized countries are possible. Definite conclusions must await the identification of the carcinogen(s) and the study of environmental factors on its production."

As the references mentioned in the report make clear, the authors made at least as exhaustive a study of the literature as did Weisburger and Reddy in their letter; they also include the papers published by these workers.

The RCP authors, however, were clearly not as convinced that this evidence was conclusive as are your correspondents. In suggesting that I am not objective in my account of the role of dietary fibre in preventing colonic cancer, Weisburger and Reddy are really saying that the RCP itself was not objective when writing its report.

JOHN YUDKIN

London NW3, UK

0028-0836/81/340666-01$01.00

© 1981 Macmillan Journals Ltd

NEWS AND VIEWS

The longest synthetic gene. . .so far

LESS than two years ago, an attempt to synthesize a human interferon gene would have literally been impossible. For it is just two years since nucleotide sequences of interferon genes were first determined by using reverse transcriptase to make DNA from messenger-RNA molecules taken from cells induced to produce interferon. Since then, similar DNA copies of mRNA have made it possible to show that there is a whole family of interferon genes in the human genome — several corresponding to leukocyte interferon (the α variety) and fewer corresponding to fibroblast interferon (called β) and the more elusive immune interferon (γ). Inevitably, the amino acid sequence of each of these interferons has not yet been worked out directly, by chemical analysis; instead, the protein structure has mostly been inferred by applying the genetic code to the sequence of nucleotides in the gene. So it is, to say the least, a technical triumph that a group assembled by ICI should now have carried through the total synthesis of a human interferon gene (see Edge *et al.*, p.756).

This gene is the largest so far synthesized, no fewer than 514 nucleotides long. Given that it has been necessary to string together the nucleotides of both the complementary DNA strands, more than 1,000 nucleotides have been assembled in a predetermined way.

The chemistry of the synthesis is more easily described than carried out. The starting point is an initial nucleotide coupled to a polyacrylamide resin. Further nucleotides are added in pairs, by means of a coupling agent in anhydrous pyridine, and the authors say that the time taken for each cycle is a mere hour and a half. At that rate, ambitious people might calculate, it should be possible to assemble 5,000 nucleotides in a year, or a whole genome in . . ., well, a few centuries. Calculations such as these show that even the modifications of DNA-synthesizing techniques worked out by the ICI group fall a long way short of assembling whole genomes.

Indeed, as now described, short sections of each strand 15 or more nucleotides in length have to be made separately so as to keep the yield of the synthesis respectably high, and then short pieces of the double helix assembled before being joined together in the complete gene. That the gene is complete has been shown by its incorporation into a plasmid, its amplification and subsequent analysis. Edge and his colleagues do not say whether their plasmids have been able to direct the synthesis of proteins — and if so, what they are — but *something* is produced, for a further report on its biological activity is promised. It would be a great surprise if the product were not α-interferon.

The gene now synthesized is not however a naturally occurring gene but an analogue, designed so as to simplify the synthesis, but also to avoid combinations of neighbouring nucleotides that may impede expression of the gene in its unnatural environment of the plasmid. It should make one of the versions of α-interferon and, with luck, should do so more efficiently. But it would be sheer good fortune if the gene now synthesized turned out to be a superabundantly efficient interferon gene. The importance of this development is not the gene itself, but the demonstration of the technique. It is a little like climbing Everest, known to be easier the second time.

But these days, it is possible to buy machines that will automatically assemble nucleotides in some predetermined sequence. What can be so special about the synthesis of this artificial gene? For one thing, the simple assembly of a score or so of nucleotides is only half the battle. Purification is a tedious and even uncertain business. Prudent people make sure that the nucleotide sequence they have made is indeed the one they set out to make by subsequent analysis. But the most important need in the construction of a largish gene is an intelligent strategy for deciding what pieces of the gene to make, and in what order to assemble them. The use of computers to spot sequences in the target gene that might cause trouble during its assembly is also obligatory.

This does not mean that the work now reported is simply a technical *tour de force*. In the past few years, synthetic pieces of genes have been widely used in several different ways — as probes for finding within some intact genome a piece of genetic material containing some particular sequence, for example. Those whose business is to insert foreign genes into plasmid DNA increasingly find it safest to resynthesize the ends of the genes back to the nearest convenient restriction site so as to ensure that nucleotides are not lost when the tails needed for ligation to the plasmid DNA are added chemically. Fashioning pieces of DNA expected to have important functions such as those acting as sites for the promotion of replication has also become fashionable and valuable. But for a group with its roots in the pharmaceutical industry, it must be especially pleasing that it is now possible to plan systematically to vary the constitution of a gene for the sake of the variations of protein structure that result. Is there just a trace of nostalgia in the statement that "classical medicinal structure–activity analysis should [now] be possible with large peptides"?

The prospect is exciting. There is already ample evidence, derived from the experience of traditional pharmacology, that natural products can usually be modified to make more efficient drugs. The changes rung on the structure of the naturally occurring antibiotics in the past several decades are a sign of that. The interferons promise to be a nightmare for pharmacologists — whatever their biological effects, unspecificity seems to be their common characteristic. So it is natural that in the past few months, there has been great interest in hybrid molecules of interferon — one end from one and the other from another. Now it should be possible to set about making more subtle modifications, not those that depend on where some restriction site within the gene happens to fall. The result should help throw light on the functions of different parts of the interferon molecule. Yet improving on naturally occurring interferon is bound to be a daunting task. While reliable biological assays of the effectiveness of molecules related to interferon are few and far between, and while the efficiency with which interferon genes embodied in plasmids yield protein product is as low as at present, it would be hoping for too much that the uncertainties that abound in the genetic engineering of medicinal interferon will quickly be cleared up.

0028-0836/81/340667-02$01.00

© 1981 Macmillan Journals Ltd

Saturn's rings

MUCH of the interest in Saturn's rings during the impending close encounter of Voyager II will centre on explanations of the structure of the ring system, involving either resonance with the major planetary satellites or the effects of smaller satellites co-moving with the rings. The excitement may obscure the argument due to Joseph E. Avron and Barry Simon (*Phys. Rev. Lett.* **46**, 1166; 1981) published earlier this year that the radial structure of the rings of Saturn may simply reflect the instability at certain radii of solutions of the full many-body problem.

The argument is simple, although incomplete. Ignoring all but Saturn and one ring particle, supposed of negligible mass, orbits with all possible periods of revolution between that of the orbit that grazes the equator of the planet and infinity are stable. No more complicated version of the problem is, however exactly soluble. The general equation of motion is, however, $\ddot{\mathbf{r}} = \mathbf{F}(\mathbf{r}(t))$ where the force-function **F**, a vector function, itself entails a solution of the many-body problem of Saturn, together with its substantial satellites and, for that matter, the Sun.

Approximations are evidently necessary, one of which is that the true force-function should be represented as the sum of forces within the equatorial plane and those perpendicular to it — whereupon it is also sensible to represent the radius vector as the sum of a vector in the plane and another perpendicular to it with the form $w(t)$. To a first approximation, supposing w to be small, the result is the need to solve a Schrodinger equation with (at best) a quasi-periodic potential. This simplest version of the problem of the motion of Saturn's ring particles can, in other words, be approximated to by the still insoluble problem of knowing what states are accessible to an electron in a quasi-periodic electric potential.

The upshot is that many periodic solutions of the simplest equations are unstable and therefore, over a substantial period of time, untenable. But which are they?

Avron and Simon did not pretend, in their paper published at the end of April, to be able to calculate the periods of the orbits that would be unstable (and thus the position of the gaps in Saturn's rings). Instead, they argued that even the simplest approximation to the many-body problem gives rise to Schrodinger-like equations whose solutions are restricted. Some orbits are possible and stable. Others are not. Avron and Simon surmise that between the rings, stable orbits should be distributed in radius as the numbers in a Cantor set — sparsely but not isolated.

Solar spin variation

from David W. Hughes

ONE of the most striking features of normal main-sequence stars is the great difference in the angular momentum per unit mass between the early and late types. There is a sharp drop at spectral type F, as shown in Fig.1. Some braking process must act preferentially on stars of low mass.

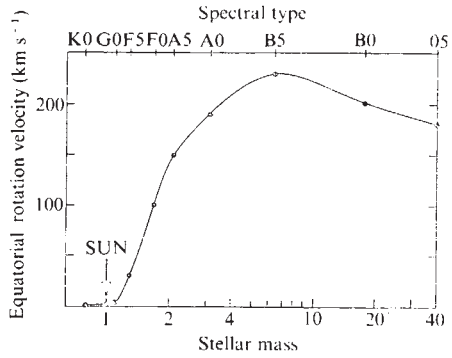


Fig.1 The equatorial rotational velocity of main-sequence stars plotted as a function of spectral type and stellar mass (measured in units of the solar mass).

Schatzman (*Ann. Astrophys.* **25**; 1, 1962) pointed out that braking would occur if the gas emitted by a star was magnetically constrained to co-rotate with the star out to distances that were large compared with the stellar radius. Under these circumstances a small amount of mass loss would yield a proportionally much greater loss of angular momentum, simply due to the effective increase in the moment of inertia of the outflowing gas. The efficiency of the braking depends on the strength of the subsurface convection which is itself closely related to the mass of the main-sequence star. This convection is also responsible for the surface activity, such as sunspots, for the corpuscular emission associated with these spots and for part of the general magnetic field of the star.

Edward H. Geyer of the Hoher List Observatory, University of Bonn, has studied the braking of solar rotation by magnetic activity and his conclusions have recently been published in *The Moon and the Planets* (**24**, 399; 1981). Geyer compares the Sun with the primary component of the eclipsing binary XY Ursae Majoris (UMa). This is a solar-type star, with similar mass but about 50 times more magnetic activity than the Sun. The binary has an orbital period of 11.5 hours, the components are close and suffer considerable tidal deformation. The XY UMa primary has a spot cycle of about 3.7 years, in comparison with the solar cycle of 11.04 years. At maximum the XY UMa primary has about 20 per cent of its surface covered in spots. XY UMa is spinning quickly — the braking that would lead to a loss of rotational angular momentum is counteracted by a replenishment from the much larger orbital angular momentum of the secondary by a tidal feedback mechanism.

The rotational angular momentum of a star on the lower main sequence determines its spot activity. As magnetic brakes have been applied to the Sun for at least the past 4.6×10^9 years, the Sun must have been spinning faster and showing more magnetic activity in the past. Geyer assumes that the loss of angular momentum takes place at the same heliographical latitudes (5° to 40° north and south) as the maximum spot activity; this maintains the differential rotation. He also assumes that the mass loss is less than a few per cent of the present mass and that the radius has remained constant. He concludes that the angular velocity and the spot activity both decay exponentially, with a mean lifetime of 1.55×10^9 years. Today the mean sunspot number, $\bar{R}$, is 51 ± 35 and the equatorial spin velocity is 2.02 km s^{-1}. The values of these quantities in the past are plotted in Fig.2.

The effects of these changes are interesting. In the past, the faster spinning Sun would be more oblate, which would perturb the orbital motion of the inner planets. The greatly enhanced sunspot activity must have influenced the evolution of life on the Earth. The increase in solar particle flux and extreme ultraviolet radiation would disturb the terrestrial magnetic field and the original atmosphere; the production rates of carbon-14 and tritium would be higher than they are today; and the atmosphere would be enriched in ^{3}He.

Geyer concludes that evolution of life and its transference from the protecting water to the exposed land were governed by the solar activity and its decay with time. □

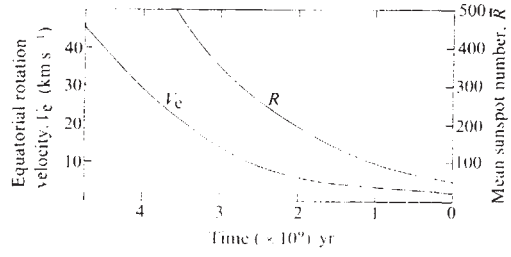


Fig.2 The decay of the solar equatorial rotational velocity, V_{3e}, and the mean sunspot activity, $\bar{R}$, as a function of time. Zero represents the present day.

David W. Hughes is in the Department of Physics, University of Sheffield.

0028-0836/81/340668-01$01.00

©1981 Macmillan Journals Ltd

Cosmochemistry and the origin of life

from Geoffrey Eglinton, Ann Henderson-Sellers and Stephen Moorbath

AT a recent NATO conference* on 'Cosmochemistry and the origin of life', Ponnamperuma (University of Maryland) summed up the overall picture by showing that there is a continuity stretching from space through to the present in terms of the formation of organic molecules and those that are found in present-day living organisms, and held that this is the natural course of events. Viola (Indiana University) started with the 'big bang' and then discussed nucleosynthesis and the abundance of the elements, particularly those essential to life. He also discussed the life cycle of stars and models of the Universe — salutary topics which should be compulsory course work for all scientists at some stage in their careers. Greenberg (University of Leiden) and Irvine (Chalmers University, Sweden) continued with details of the nature, composition and distribution of the interstellar dust and gas medium, and emphasized the enormous reservoir of organic molecules and most basic raw materials of life known to exist in interstellar space. Conference participants thought it unnecessary, at this state of our knowledge, to invoke *Panspermia* — the seeding of our planet, at any stage in its history, by life from extraterrestrial sources. In this much publicized context the conference generated neither science fiction nor science friction.

NASA's spaceship Enterprise, ably piloted by de Vincenzi (NASA Headquarters), took us on a tour of the Solar System, culminating in an account of the Mars Lander's biological experiments. Initial optimism and excitement soon gave way to the more sober realization that surface oxygen release could be better accounted for by presence of labile, active inorganic oxides in martian soil than by biological activity. Prospects for contemporary life on Mars are bleak, but perhaps in its originally airier and wetter manifestation it might have temporarily harboured the seeds of early life. The rest of the Solar System is decidedly no place for protoplasm!

The conference then rapidly came down to Earth. However improbable it seems, life and mind flourish, and some 4 billion years separate the earliest replicating organic molecules from the deliberations of the present conference. After presenting an inventory of terrestrial surface carbon, Henderson-Sellers (University of Liverpool) noted that, without its hydrosphere and biosphere, the Earth would be very similar to Venus, where carbon dioxide is not tied up in carbonate rocks. She, as well as Ponnamperuma, summarized much of the evidence that the atmosphere of the early Earth was dominated by carbon dioxide, water, nitrogen and possibly some carbon monoxide, and not by ammonia, methane and hydrogen as was once widely believed. The bulk of the hydrosphere was undoubtedly present from very early prebiotic times. Ample energy sources, including electrical discharges, ultraviolet radiation, radioactivity and volcanic/geothermal heat, were available on the early Earth to foster complex, prebiotic organic syntheses from relatively simple mixtures of volatiles and gases, with clay minerals as a likely base for the earliest biotic processes.

Moorbath (University of Oxford) showed that isotopic dates on early Precambrian sediments containing biological or molecular fossils could frequently be obtained from stratigraphically closely associated igneous rocks, and more rarely from the sediments themselves. Algal stromatolites and other simple cellular structures were known from sediments probably as old as 3.5 billion years, although in some cases there is a worrying stratigraphic gap between the sediment and the dated igneous rock. Whether the Isua metasediments of West Greenland, dated at about 3.8 billion years, contain biogenic markers is still a matter of controversy. Schidlowski (Max-Planck-Institut, Mainz) is convinced from detailed carbon isotope studies that Isua carbonaceous sediments represent a 3.8 billion year-old record of biological activity. There is a continuity of the $\delta^{13}C$ record from the present back to Isua times, and yet the Isua rocks also exhibit a marked shift in $\delta^{13}C$ which may result from their strong metamorphic overprint. Indeed, some workers claim that the $\delta^{13}C$ record of the Isua rocks has been achieved by metamorphism alone. Walters (University of Maryland), after painstakingly pyrolizing 27 Isua graphite samples, observed no trace of biogenicity, with one possible exception. The metamorphism was certainly intense enough to erase any biogenic evidence.

Ochiai (University of Maryland) pointed out that bioinorganic chemistry was essential for understanding chemical evolution. Many trace metals, such as Fe, Cu, Mo, Mn, Co, Zn, Ni, Cr and V, form complexes which catalyse organic reactions considered essential for life processes. He also discussed the problem of the major Precambrian Banded-iron Formations, ranging in age from about 3.8 to 2.2 billion years, and regarded by many as biogenic markers in the sense that their precipitation may have required oxygen released from marine organisms, in the virtual absence of atmospheric oxygen. This 'iron age' was followed at about 1.5 billion years onwards by copper sulphide deposits — the 'copper age'. Before this, prevailing redox conditions would not permit copper to appear in a dissolved, mobile state. A tentative correlation might be made between the sudden appearance of the copper deposits and the distribution of copper enzymes in present-day organisms. In this view, the copper superoxide dismutase is a later and more efficient development than the manganese and iron superoxide dismutases.

From the organic record in more recent sediments, Eglinton (University of Bristol) produced abundant evidence that molecules can retain a high degree of structural and stereochemical specificity over hundreds of millions of years if they are not exposed to raised temperatures or subject to oxidation or microbial attack. Specific biological marker compounds, or 'chemical fossils', can be interpreted in terms of the palaeoenvironment. The thermal history of a sediment can also be understood in terms of epimerization, aromatization and carbon–carbon bond cleavage reactions observed in samples with different burial histories. Of particular interest is the possibility of relating the molecular lipid record to microbial activity. Thus the archebacteria display unusual membrane lipids which may provide a clue to their activity during the Precambrian.

All participants agreed that there is an urgent need to continue the search for ancient, unmetamorphosed sediments from the earliest accessible Precambrian, so that micropalaeontologists, inorganic and organic geochemists, clay mineralogists, geologists and geochronologists can work closely together towards a progressive characterization, through geological time, of the biological record in the rocks. Recent geological work clearly shows that the earliest known (about 3.8–3.5 billion year) sedimentary and volcanic rocks characterize a depositional environment which is totally compatible with conditions now widely regarded as favourable for the earliest life on Earth. Of course, one big question was forcibly put to us by the students — "when did life begin?". The best bet was 4.0±0.1 billion years ago. The origin of life could be even older, but most people agreed that it was unlikely to be younger.

Altogether, it was a creative and original Study Institute, which looked as much into the future as into the past. Is it just conceivable that a conference on the same topic is at this moment being held somewhere else up there in the night sky?

*The NATO Advanced Study Institute was held in Maratea, Italy on 1–12 June 1980. The director of the institute was Professor Cyril Ponnamperuma.

Geoffrey Eglinton is in the School of Chemistry, University of Bristol; Ann Henderson-Sellers is in the Department of Geography, University of Liverpool; and Stephen Moorbath is in the Department of Geology, University of Oxford.

0028-0836/81/340669-01$01.00

©1981 Macmillan Journals Ltd

Holocene wiggles

from R. W. Fairbridge

THE trouble with time series relating to geophysical and climatic phenomena is that the instrumental data base is often less than 100 years long, whereas many of the suspected planetary steering mechanisms operate on quasi-periodic wavelengths in excess of 10^3 years. More precise chronostratigraphical methods are needed when looking for correlations in the area of celestial mechanics where hindcasting may often be estimated with a year-to-year accuracy, for at least 10^4 years. The customary geological dating technique in this age range, radiocarbon, commonly suffers from an accuracy of only ± 250 years or so.

Some new developments discussed at a recent meeting* promise to provide a year-to-year chronology dating back to beyond 13,000 BP, with a precision of ± 1 year. First, there is the old varve chronology of De Geer. To begin with it suffered from a 'bad press'; people did not feel much confidence in a simple thickness counting. But over the past two decades the Swedish scientists and government have considered it important enough to spend several millions of kronor proving the accuracy of the method (within the error limits of radiocarbon), and re-counting the old series. I. Cato (University of Uppsala) recently completed the count up to AD 1980 and incidentally has shown that it is the annual peak flood (1–2 days) that corresponds to the maximum thickness of the summer varve. Soon it will be possible to state the peak flood years in south and central Sweden for the past 13 millennia.

Second, the $^{18}O/^{16}O$ analyses of ice cores taken in Greenland and Antarctica are being extended and 'massaged' to generate far more precise time series than at first seemed possible. At the meeting, C. Langway (SUNY, Buffalo) described some of the techniques and W.D. Hibler (Cold Regions Research Laboratory, New Hampshire) demonstrated an approximately 20 year cycle recognized in both ice cores and tree rings since AD 530. W. Dansgaard (University of Copenhagen) recently reported[1] some of the ice-core problems and solutions. Absolute (year-to-year) counting now goes back nearly 2 millennia and is rapidly being extended. Hammer, Clausen and Dansgaard[2] have described an ingenious *p*H-measuring device that records the acid layers in the ice cores that develop as a result of volcanic gas 'rain-out'. As Clausen (University of Copenhagen) demonstrated at the meeting, it is now possible to identify many historically known volcanic eruptions and thus independently confirm the layer counts. When combined with the oxygen isotopes, this gives a climatic reading.

The third category of Holocene wiggly-line time series concerns dendrochronology. In contrast to varves and ice cores that reflect regional climatic trends, tree rings initially reflect only individual trees, but with regional studies and sophisticated analysis this material also furnishes valuable climatic series. In addition, and most importantly, the tree rings provide a fourth wiggly line — the variations in the radiocarbon flux — from which come the cosmic ray and solar wind fluctuations. We now have a picture of solar activity that takes us back nearly 8,000 years[3], and this limit can be extended to beyond 13,000 BP as soon as some floating chronologies can be tied in.

The problem of floating time series is a speciality of D. Schove (St David's College, UK). There are numerous varve and dendrochronological series from Europe, North America, South America and Asia that have not been tied in to a known datum. If these long runs can be spliced together we shall get a year-by-year record of climatic events (local, regional and, we hope, global) going back nearly 14,000 years. Schove works on the principle that most of the wiggles represent local climatic 'noise' and may have little global significance, but certain key years disclose very remarkable 'spikes' which recur in distinctive patterns. The methodology works very well with tree rings, varves and, on a longer time scale, geomagnetic reversals.

Simplistic correlation attempts with the 11 year sunspot 'cycle' have often run aground, but analyses of phase changes and interactions with other cyclic effects can be rewarding, as reported by J. Harlin[4]. Schove[5] has considered the fact that the quasibiennial cycle (26 months), when superimposed on a powerful annual signal, may eventually lead to a 3 year jump. Inasmuch as the sunspot maxima fluctuate markedly, the combined effects result in distinctive patterns for any given century. Although the astrophysical and meteorological mechanisms are still shrouded in mystery, the pattern system seems to work, providing the approximate time span is known — and this link is provided by radiocarbon and other geological techniques. Particularly important time series that may soon be tied in by these 'teleconnections' are the Finnish varves, described by M. Saarnisto (Oulu University, Finland), and those of southern Germany by J. Merkt (University of Hannover). Schove has proposed teleconnections in the mid-Holocene between America (Bristlecone), Ireland, the Crimea and Anatolia[6].

Hints about the long-term variables were to be found in the palaeomagnetic records displayed at the conference by K. Creer (University of Edinburgh). Systematic coring of undisturbed lake beds from different parts of the world is now beginning to bear fruit. Both inclination and declination show systematic fluctuations with a quasi-periodicity of the order of 1,000–3,000 years. A variable rate of westerly secular drift introduces a longitudinal lag in dates, and at the same time warns of the variability in the outer core dynamics.

Inasmuch as both the geomagnetic field and the solar wind (a magnetized plasma

*A conference on 'Holocene Correlation' was held on May 1, 1981 at the Warburg Institute, University College London, under the auspices of the UNESCO-supported International Geological Correlation Programme no.158. The meeting was jointly chaired by Dr Derek Schove and Professor Rhodes W. Fairbridge.

R. W. Fairbridge is Professor in the Department of Geological Sciences, University of Columbia.

100 years ago

NOTES FROM THE MALAY ARCHIPELAGO

A correspondent in Java sends us the following:-

The cattle plague has been raging in the west end of Java, Bantam, the Preanger, and Batavia residences — during the west monsoon (now finishing) with redoubled vigour. It has now abated a little (after four years it may well do so, from want of victims) in these parts, but is extending eastward, its appearance in Krawang being the most alarming. The authorities have decided upon making a double-fence right across Java at its narrowest part. This means a line from somewhere about Cheribon due south. In the interval — a considerable one — between the two fences, no cattle will be allowed to pass or exist.

There is a bird (native name Jallak) which follows the buffaloes about and perches on their backs. Query, can this bird have anything to do with the spreading of the plague? If so I don't see what Government can do. They can't fence him out.

In all the parts where the cattle-plague has raged the most awful fevers have been the result amongst the native population. In Bantam alone 50,000 died in 1880. In the Preanger and Batavia the death-rate was also very high. There is no doubt whatever that this is due to the imperfect interment of the carcases. The Government says it is due to the wet season; but this is a lame excuse, for why is there no fever elsewhere? In the wet season it is, of course, worse, for the heavy rains cause more miasma.

From *Nature* **24**, 25 August, 399, 1881.

0028-0836/81/340670-02$01.00

©1981 Macmillan Journals Ltd

flow) modulate the incoming galactic cosmic rays, the radiocarbon flux, monitored by the tree rings, can now be related in part to core dynamics. A new monitoring isotope, beryllium-10, was discussed at the meeting by G.M. Raisbeck (Orsay, France). This isotope is also formed in the atmosphere by cosmic ray collisions and within a year deposited on the Earth's surface. It can be measured most conveniently in ice cores, thereby providing not only a way of cross-checking the ^{14}C data, but also of extending the cosmic radiation record back for perhaps up to 10^5 years.

R.W. Fairbridge (Columbia University) displayed colour slides of the field evidence illustrating how the cyclicity of high-latitude storminess cycles on the Hudson Bay seems to reflect a 45 year 'Double Hale' solar cyclicity[7]. He then reviewed new unpublished works that demonstrate solar–planetary periodicities from the Permian period, 250–280 million years ago: first, from an anhydrite series in New Mexico (R.Y. Anderson, University of New Mexico); second, from an evaporite series in the Ural Mountains, USSR (I. Vulis); and third, from a varve series in Brazil (Rocha-Campos). The astonishing thing is that they all display what seem to be the same periodicities that occur in the Holocene, which, if nothing else, would appear to dash the hopes of velikovskian catastrophists.

1. Dansgaard *Nature, News & Views* **290**, 360 (1981).
2. Hammer, Clausen & Dansgaard *Nature* **288**, 230 (1980).
3. Eddy *Scient. Am.* (1977).
4. Harlin *Phys. Geogr.* **1**(2), 108.
5. Schove *Weather* **26**, 201 (1971).
6. Schove *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **25**, 209 (1978); in *Moraines and Varves* (ed. Schlutter; Balkema, 1979).
7. Fairbridge & Hillaire-Marcel *Nature* **268**, 413 (1977).

Pre-equilibrium processes in nuclear reactions

from P.E. Hodgson

THE processes that take place when a fast nucleon (10–100 MeV) interacts with a nucleus are quite well understood, but there are many detailed features that require further study. The incident nucleon first collides with a single nucleon or perhaps a cluster of nucleons in the target nucleus. The products of this collision have a high probability of escaping from the nucleus, particularly if the collision takes place on the periphery. This initial interaction takes place rapidly, in the time it takes for the incident particle to traverse the target nucleus, typically about 10^{-22} seconds.

The particles from this first collision that do not escape from the nucleus undergo further collisions until all their energy is transferred to the compound system. The energy is shared among the nucleons until eventually, by statistical fluctuation, a nucleon or group of nucleons receives enough energy to escape; this continues until the energy falls below the particle emission threshold. The residual nucleus eventually attains its ground state by γ and β decay. The statistical processes by which the compound nucleus loses its energy take much longer than the direct process, typically about 10^{-16} seconds.

The direct and compound processes are quite well understood theoretically, so the corresponding cross-sections can be calculated with some confidence. The simplest direct process, elastic scattering, can be calculated by the optical model, and inelastic scattering by the coupled channels or distorted wave formalism. Nucleon transfer reactions can be treated by the distorted wave Born approximation, and the probabilities of emission of various particles from the decay of the compound nucleus can be calculated using the Hauser–Feshbach theory with the transmission probabilities obtained from optical model analyses of elastic scattering.

The direct and compound processes usually both contribute to each reaction and the total cross-section may be obtained by adding them incoherently. This may be compared directly with the measured cross-sections, after the latter have been energy-averaged to remove the fluctuation in the compound nucleus cross-section.

Such calculations are generally fairly successful at lower energies, but at higher energies additional processes can occur and have to be accounted for. In particular, it is possible for particles to be emitted after the direct process but before the nucleus has attained statistical equilibrium. These pre-equilibrium or pre-compound particles can be recognized experimentally and their cross-sections have been calculated using several different models.

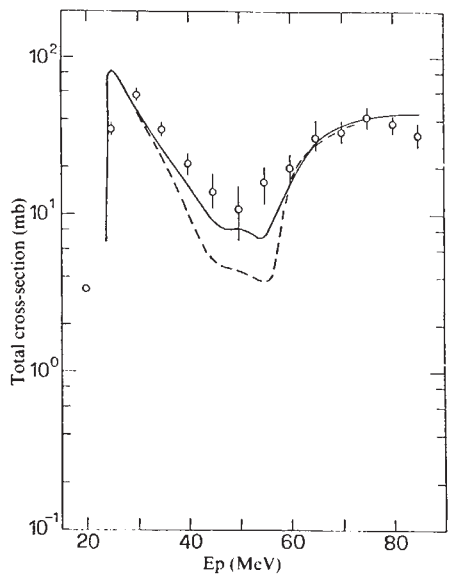


Fig. 1 The excitation function for the $^{50}Ti(p,2p2n)\ ^{47}Sc$ reaction compared with exciton model calculations. The dashed line shows the result of a calculation without the pre-equilibrium particles and is clearly inconsistent with the data. From ref. 3.

P.E. Hodgson is in the Department of Nuclear Physics, University of Oxford.

The particles emitted in the direct, pre-equilibrium and compound stages are characterized by their differing energy and angular distributions. Direct particles are emitted predominantly in the same direction as the incident particle, with similar energies, whereas compound nucleus particles are emitted in all directions, equal numbers forwards and backwards, and when the final states merge into a continuum have a characteristic maxwellian energy distribution. Pre-equilibrium particles tend to be emitted forwards, and are generally more energetic than those from the compound nucleus.

To calculate the cross-section for pre-equilibrium emission, the interaction process is considered to take place in several stages. The first stage can lead to the emission of direct particles, or to the excitation of a particle-hole pair. In the second stage further interactions can occur, which can lead to the excitation of further particle-hole pairs, and so on until finally the compound nucleus is formed. At each stage in the excitation process it is possible for pre-equilibrium particles to be emitted.

This process has been formulated mathematically, enabling the pre-equilibrium cross-sections to be calculated. One type of theory due to Blann[1], Griffin[2], Gadioli[3] and others gives semi-classical expressions for the cross-sections in terms of the basic transition matrix elements that express the probability of going from one stage to the next. Such theories have been used to calculate the cross-sections of many different reactions and a typical comparison with experiment is shown in Fig. 1. In this work the cross-section is determined radiochemically, so it is not possible to distinguish between the (p, α) and the (p, 2p2n) reactions. The former reaction is favoured energetically and is responsible for the first peak, and the latter for the second. The two curves are calculated with and without the pre-equilibrium contribution, and comparison with the experi-

0028-0836/81/340671-02$01.00
©1981 Macmillan Journals Ltd

mental data shows the presence of these particles.

More recently, a quantum-mechanical theory has been developed by Feshbach, Kerman and Koonin[4], and numerical calculations made by Colli-Milazzo and collaborators[5]. This gives both the angular and energy distributions of the emitted particles, and extensive comparisons with the experimental data for (p,n) reactions from 25 to 45 MeV on nuclei from calcium to tin show good overall agreement. An example of such a comparison is given in Fig.2, showing the contributions of the one-step (direct) and various multi-step (pre-equilibrium) processes.

Calculations of the contributions of multi-step processes to proton inelastic scattering have been made by Tamura and Udagawa[6] using an extended version of their distorted wave formalism. This theory is also able to give a good overall account of the measured cross-sections.

There is still much work to be done to justify and establish the accuracy of some of the assumptions and approximations made in these calculations, and it will be important to test the different theories by using them to analyse the same reactions. □

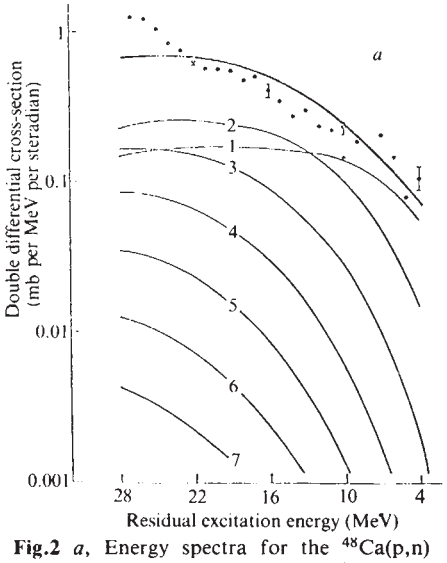
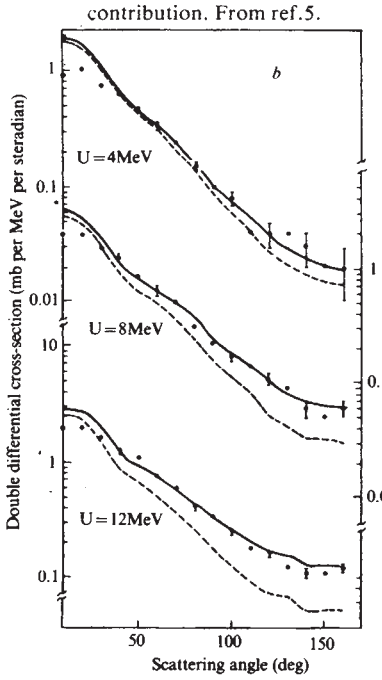


Fig.2 *a*, Energy spectra for the ^{48}Ca(p,n) reaction at 45 MeV at 90° as a function of the energy of the residual nucleus. The individual curves show the contributions of the various multistep processes. The excess of particles at higher excitation energies, corresponding to lower emission energies, is attributed to multiparticle emission. *b*, Differential cross-section for the ^{90}Zr(p,n) reaction at 45 MeV showing the single-step (dashed line) contribution. From ref.5.



1. Blann, M. *A. Rev. nuclear Sci.* **25**, 123 (1975).
2. Griffin, J.J. *Phys. Rev. Lett.* **17**, 478 (1976); *Phys. Lett.* **24B**, 5 (1967).
3. Gadioli, E., Gadioli, Erba, E., Hogan, J.J. & Burns, K.I. *Phys. Rev.* (1981).
4. Feshbach, H., Kerman, A. & Koonin, S. *Ann. Phys.* **125**, 429 (1980).
5. Colli-Milazzo, L., Bonetti, R., Camnasio, M. & Hodgson, P.E. *Phys. Rev.* (1981).
6. Tamura & Udagawa, T. *Phys. Lett.* **71B**, 273 (1977); **78B**, 189 (1978).

Mammal tracks on mountain slopes

from Robert M. May

ONE of the many differences between the Victorian Age and our own is written in the paths in the English Lake District. Today's path from Wasdale Head to Sty Head follows the direct line (and the steepest slope) across the shoulder of Great Gable; few use the nineteenth century path that zig-zags in a more leisurely and scenically varied way along the side of Lingmell Beck. From the head of Langdale one can ascend to Angle Tarn directly by the stoney boredom of Rossett Gill, or more slowly via the sweeping switchbacks of the almost vanished pony track.

Faced with these choices, the behavioural ecologist may be prompted to ask which paths are energetically optimal. Some pertinent laboratory and theoretical studies are due to Taylor, Caldwell and Rowntree (*Science* **178**; 1096, 1972). These authors began by observing that the energy needed to elevate an animal to a given height is proportional to its body weight, whereas the animal's power output available to perform this task tends to be roughly proportional to its surface area; hence the energy costs of moving uphill (compared with moving at the same speed on level ground) are relatively greater for larger animals. Turning to the laboratory, Taylor *et al.* substantiated this analysis by showing that movement up a 15° incline at 2 kmh^{-1}, compared with running on level ground at the same speed, resulted in a 40 per cent increase in oxygen consumption for a chimpanzee, but only 1–2 per cent for a mouse. Extrapolating their experimental results, they estimated that running vertically at 2 kmh^{-1} would increase the oxygen consumption of a 30 g mouse by 23 per cent, of an 18 kg chimpanzee by 190 per cent, and of a 1,000 kg horse by 630 per cent.

This work helps explain why squirrels can run up trees at much the same speed as they run from tree to tree, and why fell runners are lean. But it sheds no direct light on whether largish animals should ascend hills directly, or obliquely in zig-zags. Indeed, an analysis along the lines of Taylor *et al.* suggests that the energy involved (in addition to baseline metabolic costs) in getting an animal of mass m to the top of a mountain of height h is simply mgh, independent of the angle of attack. A more gradual ascent, at an angle of ascent θ relative to the horizontal, involves a relatively smaller cost of moving up the slope (by a factor $\sin\theta$), but takes longer to reach the top (by a factor $1/\sin\theta$); the two effects cancel each other out.

Reichman and Aitchison (*Am. Nat.* **117**; 416, 1981) have recently made a most interesting attempt to find out how animals actually do ascend or traverse hillsides in nature. To this end, they measured trails left in the snow on hillsides in the south-western United States, by various mammals. By choosing fresh snow that covered existing trails, these authors hoped to find trails that had been made at the discretion of the individual animals; records were compiled only for situations where the animal appeared to be moving freely up a slope, rather than wandering from one feeding spot to another, or fleeing pursuit. Each mammal was identified by its footprints, and the average weight for that species assigned to it. In this way, Reichman and Aitchison compiled data for some 130 individual animals, of 22 different weights ranging from a 0.025 kg mouse to a 680 kg buffalo.

Reichman and Aitchison defined the 'trail angle' to be the angle between the actual trail on the hillside and the horizontal; thus an animal proceeding directly up the mountain at the angle of steepest ascent has a trail angle of 90°, while an animal moving around the hillside at a constant elevation has a trail angle of 0°. This trail angle has a slight, but statistically very significant, propensity to be smaller for heavier animals. The negative correlation is even more striking if one

Robert M. May is Class of 1877 Professor of Zoology at Princeton University.

0028-0836/81/340672-02$01.00

©1981 Macmillan Journals Ltd

plots the maximum trail angle observed for any one body size against the weight of the animal. As Reichman and Aitchison observe, "it seems intuitive that the trail angle used by animals should decrease with increased body size, and the field data support this outcome".

A clearer test of the basic ideas, however, is provided by examining the correlation between trail angle and the slope of the mountain itself. The analysis outlined above suggests that, to a first approximation, the energy costs are the same whether one ascends obliquely or directly. This would lead to the expectation that the trail angle is uncorrelated with the slope of the mountain. Alternatively, for animals in a hurry to get to their destination, we might often expect a direct ascent, and hence trail angles of 90° irrespective of the mountain slope. This latter expectation may be modified for large animals on steep slopes, where the metabolic costs of sustaining uphill motion may become prohibitively large; that is, there may be a maximum slope, $\theta_{max}(m)$, beyond which an animal of mass m cannot sustain uphill motion. Such a θ_{max} will decrease with increasing body weight; on an almost vertical hillside, mice and squirrels may scamper up directly (trail angle of 90°), but heavier animals are likely to ascend obliquely at a trail angle of about θ_{max}. Of course, questions of local topography and , above all, the purpose of the journey would complicate all this.

On this basis, I would expect that, when Reichman and Aitchison's data (embracing all body weights) were plotted in the form of trail angle versus mountain slope, there would either be no correlation at all, or else a propensity for trail angles to be random at low mountain slopes, and for steep trail angles to be rarer at high mountain slopes. In fact, such a plot reveals a puzzling tendency for trail angle to be approximately proportional to mountain slope. Remarkably, essentially no animals — not even the mice — are ascending shallow hills directly (at trail angles of 90°).

It seems likely that the animals making these tracks were, by and large, simply wandering across or skirting the shallow-sloped mountains. (Though why they should be doing so almost exclusively at low trail angles remains a mystery to me.) On the steeper hillsides the tracks may be telling us more about energetics of moving up a slope, but even here it could be that the geometry of the tracks depends mainly on why the animal was going from A to B, rather than on anything else.

In short, the study by Reichman and Aitchison is a very original and thought-provoking one, but I think the next steps will need to include consideration of the behavioural ecology of the animals in question. If we seek to plot our paths through the hills according to principles of optimality rather than sheer delight, the current literature remains unhelpful. □

HLA genes, immunoglobulin genes and human disease

from Ken Welsh

THE human lymphocyte antigens (HLA) were recently described by one of my colleagues as simple equivalents of the ABO system found on all cells except red cells. In fact the HLA system comprises a group of several hundred genes located on the short arm of chromosome 6 but the only gene products of this region well recognized in the human are HLA-A,-B,-C,-D and -DR and complement components C2, C4 and Bf. All these are polymorphic to a greater or lesser extent and a range of products of this region is expressed on most lymphoid cells.

Almost ten years ago at an International Transplantation meeting in San Francisco, participants were given large badges inscribed with the words "Are you my HLA type?". At that time the close association between HLA-B27 and ankylosing spondylitis (AS) had just been rumoured[1,2] and prevailing evidence suggested that good HLA matching of donor and recipient would solve the major problems of transplant rejection. The badge implied not only hope for transplantation but also that by suitable selection of partner, major diseases could be avoided in the next generation. Now, although hyperacute rejection has been virtually eliminated, we are no closer to predicting acute or chronic rejection on the basis of HLA type and the closest HLA and disease association is still that between HLA-B27 and AS. But the years of consolidation and continual searching are beginning to be rewarded; at the recent European Dialysis and Transplantation meeting in Paris we heard how revisitation of red cell types and skin testing techniques coupled with precise HLA cross-matching and HLA-DR typing have again given us hope for success in transplantation.

On the question of HLA and disease, a paper published in this week's issue of *Nature* (p.768) by Uno and colleagues may provide the necessary break in the flood wall and enable us accurately to identify individuals at risk from certain major diseases. Several cracks had already appeared in the wall, so a break is not unexpected.

The first came with the finding that a combination of HLA type and acetylator status can identify with very high reliability those women who will develop autoimmune complications if treated for hypertension with the drug Hydrallazine[3]. HLA-DR4-positive, slow acetylator females always develop the complication; DR4-positive, slow acetylator males will escape unless given high doses of the drug. About 8 per cent of hypertensive patients treated with Hydrallazine develop a lupus-like complication. In the normal population about a quarter are DR4 positive, and half are females and half are slow acetylators, giving us an expected complication rate of about 6 per cent. Occasional DR4-positive males on high drug doses, DR4-positive rapid acetylator females and DR4-negative slow acetylator females develop the complication and account for the small discrepancy. This discovery shows how a combination of apparently unrelated genetic 'markers' can be used for accurate identification of a population at risk.

Two other recent discoveries bear directly on the work of Uno *et al.* The first was that the relative risk of developing chronic active hepatitis was 11.7 for those people having HLA -B8/DR3 and 2.3 for those having a particular immunoglobulin allotype[4] — a genetic marker expressed on antibodies. As in the Hydrallazine lupus story, a combination of the two markers is associated with a significantly greater risk, in this case 40-fold greater, than if neither marker is present. The second discovery[5] was that relapse in hyperthyroid Graves' disease can be predicted by a combined analysis of HLA -DR3 status and the level of thyrotrophin receptor antibody at the time of drug withdrawal. The prediction in this case was accurate in 62 out of 65 patients.

An alternative approach to identifying individuals at risk would be to use a combination of genetic marker and family studies. For example, if incidence of a disease is 1 in a 100 in the population and the disease is associated with HLA-B27 (which is present in about 10 per cent of the population), then even if all the people who get the disease are B27 positive it still means that only 1 in 10 B27-positive people are at risk. However, if the disease is already present in a family, this risk rises very dramatically for B27 family members.

Uno and co-workers have combined these two approaches in their study of Japanese families having more than two first-degree relatives affected by Grave's disease. In the data presented, 32 out of 33 children suffering from this disease carried a particular HLA and IgG-linked gene combination. The single child having only the IgG-linked susceptibility could be explained by a recombination between the HLA marker gene and a hypothetical HLA gene actually responsible for the disease. A more complete HLA typing involving SB

Ken Welsh is in the Tissue Typing Laboratory at Guy's Hospital, London.

0028-0836/81/340673-02$01.00

©1981 Macmillan Journals Ltd

antigens, complement allotyping and additional enzyme markers would clarify the point.

Although in the face of the evidence presented one tends to think of tissue typing at birth (to predict the occurrence of autoimmune disease) or perhaps even before one goes to the computer dating service, there are practical scientific reasons for being cautious. For example, Uno *et al.* selected only 15 of the 30 families studied for inclusion without saying how or why this selection was made. Second, IgG allotype frequencies are very different in Caucasoid and Japanese populations. Third, Graves' disease is very strongly associated with B8/DR3 in Caucasoids but with DR5 and DR8 in Japanese populations. A study of Japanese families living in America might show whether these differences are due to a different environmental factor interacting with different IgG or HLA haplotypes in the two populations or to putative susceptibility genes going with different allotypes in the two populations.

Theoretically the reasons for expecting HLA and immunoglobulin-gene linked associations with immune response in general and autoimmune disease in particular are overwhelming: HLA-DR antigens (or antigens in the same chromosome area) are necessary for antigen handling and presentation by a lymphoid cell subset; markers in this region (by analogy with the mouse) are important for interaction of T cells during a response; immunoglobulin genes are also involved in T-cell recognition and control; HLA-A,-B and -C antigens are important at the effector arm of the cellular response; and finally, C2, C4, Bf and immunoglobulin are involved at the effector arm of the humoral response. Thus, HLA and immunoglobulin genes are active throughout the immune response from recognition to control and it is slight differences in an individual allele's ability to function, or more exactly a summation of slight differences, which we are measuring as disease susceptibility. Theoretically therefore we expect that use of HLA and immunoglobulin allotyping data together with other genetic markers and environmental factors should allow autoimmune diseases to be predicted exactly. However, there is still a considerable amount of analysis of both HLA-region genes and IgG-region genes to be done in order to achieve this goal in the general population.

In Japanese families, the occurrence of Graves' disease can be exactly predicted on the basis of HLA and immunoglobulin allotypes, but it is too early to start wearing "Are you my HLA type?" badges outside Japan. Nevertheless, elimination of autoimmune diseases by genetic counselling is now much closer to being realized: if two independent susceptibility genes are necessary for disease onset, then it is possible to ensure that the genes will not occur together in the next generation. □

1. Brewerton, D.A. *et al. Lancet* i, 904 (1973).
2. Schlosstein, L., Terasaki, P.I. & Bluestone, R. *New Engl. J. Med.* **288**, 704 (19737.
3. Batchelor J.R. *et al. Lancet* i, 1107 (1980).
4. Whittingham, S. *et al. Clin. exp. Immun.* **43**, 80 (1981).
5. McGregor, A.M. *et al. Lancet* i, 1101 (1980).

Positronium and positronium ions

from T.C. Griffith

POSITRONS and positronium, the bound state of an electron and a positron, have in recent years been used in a rapidly increasing range of experiments in atomic physics, solid-state and surface physics and positronium chemistry. Positronium has also played an important part in fundamental experiments designed to check the validity of certain aspects of the theory of quantum electrodynamics (QED). It is therefore pertinent that the recent detection of a new entity, the positronium negative ion, Ps^- ($e^-e^+e^-$), initially predicted to exist as a bound state by Wheeler[1] in 1946, should be greeted as an important historical landmark which could greatly extend the activities of positron physicists. The careful experiment performed by Allen Mills[2] at the Bell Laboratories, New Jersey clearly demonstrates that Ps^- can be produced in the laboratory under controlled conditions.

Ps^- is the analogue of H^- ions produced by slow proton bombardment of thin foils and the method used by Mills incorporates the same ideas. In his experiment, a nearly monoenergetic beam of low-energy positrons, guided by a longitudinal magnetic field, strikes a thin film of carbon in ultra-high-vacuum conditions. The energy of the positrons is adjusted such that the beam is partially transmitted and application of a suitable potential to a grid located at the exit side of the carbon film will reflect the transmitted positrons back to the carbon whilst accelerating the weak beam of Ps^- into a field-free drift region beyond the grid. A lithium-drifted germanium detector is used to measure the energy spectrum of the resulting annihilation photons and a γ-ray peak is found whose energy depends on the accelerating potential applied to the grid. It was demonstrated that this peak can be attributed to a Doppler-shifted annihilation γ resulting from the 2γ decay of Ps^-.

As well as being of interest in its own right, Ps^- is also important because it might be used to produce beams of positronium of well defined energies. This could be achieved by accelerating the Ps^- to a given energy and then photoionizing it to give positronium which could, perhaps, be used in scattering experiments. Another suggested application is to use the annihilation of Ps^- in flight to provide a tunable source of γ rays of precisely known relative energy shift. The intrinsic properties of the Ps^- such as its lifetime ($\sim$ 2 nanoseconds), its 2γ angular correlation, its ($1\gamma/2\gamma$) branching ratio, its magnetic moment and measurement of its photoionization cross-section are obvious candidates for extensive studies.

The glamour of detecting the Ps^- ions should not be allowed to overshadow the deeper significance of this experiment. The major triumph underlying this work is the dramatic technical advances associated with the production of intense beams of slow positrons of well defined energies achieved at Bell Laboratories and at Brandeis University by Karl Canter and co-workers[3]. Technology has now reached such a level that within a few years low-energy positron beams of intensities of 10^9 positrons per second, or even several orders of magnitude higher, can be envisaged. Positron emitters created *in situ* in a nuclear reactor followed by carefully designed moderators and beam transport systems may be used for this purpose. If allowed to strike clean metal surfaces under ultra-high-vacuum conditions these positrons can be converted, with almost 100 per cent efficiency, into low-energy positronium atoms. We might therefore expect intense and highly localized sources of positronium atoms which can be used to study the excited states of positronium using laser beams in experiments, similar to those performed for atomic hydrogen, that will provide further high-precision tests of QED.

Even more exotic possibilities exist. Quantum chromodynamics predict that orthopositronium can decay into a new elementary particle, the axion, and a γ ray. The axion must have a mass of less than 1 MeV for this to be possible and if this were the case, then an intense positronium source from a low-energy positron beam could well find application in a search for the axion. □

T.C. Griffith is in the Department of Physics and Astronomy at University College London.

1. Wheeler, J.A. *Ann. N.Y. Acad. Sci.* **48**, 219 (1946).
2. Mills, A.P. Jr *Phys. Rev. Lett.* **46**, 717 (1981).
3. Canter, K.F. Review to be published in *Can. J. Phys.*

0028-0836/81/340674-01$01.00

©1981 Macmillan Journals Ltd

Two Voyagers to Saturn

The pages which follow may be taken as a celebration of the impending encounter of Voyager 2 with Saturn and are a pointer to the richness of data likely to be gathered in the next few days and weeks. Although there may be some to whom the work of these two remarkable spacecraft will be proof of how even more remarkable would have been the Grand Tour of the Solar System planned in the 1960s but executed only in the form of the two Voyagers, to most people the encounters with Jupiter and Saturn will seem exciting enough for the time being. Yet there are Uranus and Neptune to come.

This group of scientific articles includes some of the first detailed attempts to make sense of last year's Voyager 1 observations of Saturn—and Nature *acknowledges its debt to Dr G. E. Hunt of University College, London for having helped to recruit these articles and to give shape to the ground they cover. The scope is necessarily restricted, for there is hardly a branch of planetary astronomy which has not been changed in some way by the data from Voyager 1—and which is not about to be changed again.*

How Voyager 2 has been reprogrammed

from Edward C. Stone

IN the past two years, the Pioneer 11[1,2] and the Voyager 1[3,4] encounters with Saturn have revealed a remarkable richness and diversity of physical phenomena. As a result of these discoveries and of continued Earth-based studies, major changes have been made in the observations planned for Voyager 2, providing additional opportunities for discovery and understanding of the Saturn system. These opportunities are further enhanced by the significantly different path that Voyager 2 must take through the system, passing 101,000 km above Saturn's cloud tops on 26 August as it continues on to Uranus. Of necessity, the most exciting discovery will be unexpected; but some of the opportunities for new observations are described below for each of the four major areas of investigation—the atmosphere, rings, satellites and magnetosphere.

Atmosphere

The study of the dynamics of Saturn's atmosphere is of interest because Saturn, like Jupiter, is a giant rapidly-rotating body composed mainly of hydrogen and helium, encircled by bands of clouds of frozen ammonia with large-scale wind patterns distinctly different from those of the Earth's atmosphere. On Jupiter, there are opposing eastward and westward jet streams at the edges of adjacent cloud bands, with a maximum eastward equatorial velocity of $\sim 450\ \mathrm{km\ h^{-1}}$. On Saturn, Voyager 1 found equatorial winds four times faster and found that the pattern of opposing jet streams occurred only above ~35° latitude.

Voyager 2 will not only yield a month-long time-lapse movie of the wind patterns, but will also provide a high-resolution view of the pattern of opposing jet streams in the north polar region. The winds in the atmosphere above the clouds exhibit a different pattern which will be determined from a high-resolution thermal map of the upper atmosphere. Other studies include a deep probing of the temperature and pressure of the atmosphere by the radio beam as the spacecraft disappears behind the planet, where images of the dark side of Saturn will be taken to search for lightning. Since Voyager 1 discovered UV auroral emissions at 80° latitude and similar emissions near the equator, Voyager 2 will also perform a systematic latitude search for such emissions.

Rings

The rings, the second major area of study, presented most of the surprises from Voyager 1, and the Voyager 2 observations programme has, as a result, been almost completely revised and enlarged. The Voyager 2 observations will be further improved by the increased solar illumination of the north face of the rings as the Saturn year continues towards northern summer and by a much closer approach to the rings than on Voyager 1. Unlike both Voyager 1 and Pioneer 11, Voyager 2 will view the lighted face of the rings at closest approach.

One of the Voyager 1 discoveries was that the bright broad B-ring appears to consist of hundreds of ringlets separated by narrow gaps, each one of which may be cleared by small, kilometre-sized satellites. The Voyager 2 photopolarimeter will observe the star Delta Scorpii through the shadowed region of the rings for 2 h 18 min, during which time the star will move radially outward behind the rings, its apparent intensity varying as it moves from behind a ringlet into the adjacent gap. This stellar occultation should determine the location, width and spacing of the ringlets with a precision of 100–300 m. A second, shorter stellar occultation observation of Beta Tauri will provide five times more precise measurements of the braided F-ring and the outer edge of the main ring system. Near closest approach, the highest-resolution images should reveal any structure down to the sub-kilometre scale.

The dynamics of the braided F-ring will be studied in a sequence of observations keyed to assessing the importance of various physical processes in its formation. The fine submicrometre-sized material in the braids is radially confined by two shepherding satellites (1980 S26 and 1980 S27) on either side of the F-ring. Images of the F-ring near these two satellites may indicate whether they are perturbing the motion of the ring particles in such a way as to cause the braided appearance while repeated images of the ring should indicate whether the braided structure is stable or varies on a short time scale. Images of the portion of the ring emerging from Saturn's shadow may provide clues to the influence of electrostatic charging of the small ring particles on the braided structure, since the electrostatic charge on the particles will probably change with the cessation of photoelectron emission normally occurring in the sunlight.

The 'spokes' in the B-ring are also not understood, although they appear to form sporadically in Saturn's shadow and then almost to dissipate before the completion of a single orbit about Saturn. Like the F-ring braids, the spokes consist of fine particles, possibly elevated in a cloud above the ring plane. As Voyager 2 crosses the ring plane, three images will be taken of the rings edge-on, in search for any elevated material. Before closest approach, a 13.5-h time-lapse movie will be taken to study the dynamical properties of individual spokes as they are distorted by the greater angular velocity of the spoke material nearer Saturn.

Edward C. Stone is the Chief Scientist of the Voyager Project at the Jet Propulsion Laboratory and Professor of Physics at the California Institute of Technology, Pasadena.

0028–0836/81/340675—02$01.00

© 1981 Macmillan Journals Ltd

Table 1 Voyager 2 closest approach distances to Saturn's satellites (23–26 August 1981)

Name	Diameter (km)	Orbital radius (R_S)*	Closest approach distance (km)†
1980 S28	40×20	2.28	287,170
1980 S27	220	2.31	246,590
1980 S26	200	2.35	107,000
1980 S3	90×40	2.51	147,010
1980 S1	100×90	2.51	222,760
Mimas	390	3.08	309,990
Enceladus	500	3.95	87,140
1980 S25	30–40	4.88	284,400
Tethys	1,050	4.88	93,000
1980 S13	30–40	4.88	153,520
1980 S6	~160	6.26	318,200
Dione	1,120	6.26	502,250
Rhea	1,530	8.74	645,280
Titan	5,140	20.25	665,960
Hyperion	290	24.6	470,840
Iapetus	1,140	59	909,070
Phoebe	~160	215	2,075,640

* Mean distance of the satellite from the centre of Saturn in units of Saturn radii (1 R_S = 60,330 km).
† From a distance of 100,000 km, the geometrical resolution of the imaging system is 2 km per line pair.

The closer approach to the rings also makes possible further study of the electrostatic discharges detected by Voyager 1. Although the origin of the discharges has not been directly determined, they occur episodically every 10 h 10 min, suggesting an association with a satellite with this orbital period embedded within the B-ring. Additional high time-resolution samples of electromagnetic wave activity will be acquired over the 10–40-MHz frequency range, including one when the spacecraft crosses the ring plane.

Satellites

Another major area of study is the saturnian satellites, of which 17 are now known. Of these, only Titan is a major satellite, similar in size to the planet Mercury. Titan, with a dense atmosphere which includes nitrogen and methane in which organic photochemistry is occurring, was a primary target for Voyager 1. Voyager 2 will observe Titan from a much greater distance (see Table 1), returning information on the light-scattering properties of the photochemical haze and possibly on the winds.

Seven of the satellites are intermediate in size, composed mainly of water ice, with diameters ranging from about 300 to 1,500 km. Of these, Voyager 1 observed Mimas, Dione and Rhea with high enough resolution to show that their icy surfaces are heavily cratered from impacts, as was expected. Unexpected, however, were regional systems of fractures and sparsely cratered plains which are evidence of significant crustal evolution. Voyager 2 will extend the high-resolution coverage to Tethys, which has a ~750-km-long valley, and to Enceladus, which is highly reflective and appears to be relatively devoid of impact craters. The absence of craters on Enceladus would require a source of energy large enough to cause significant modification of the surface. To investigate Enceladus's thermal properties, the rate of cooling of the satellite as it moves into Saturn's shadow will be determined from IR temperature measurements.

Less is known about Hyperion and Iapetus, the other two intermediate satellites. Voyager 2 will fly closer to both than did Voyager 1, returning images with enough resolution to determine whether there are large valleys as on Tethys or whether there is any other evidence of crustal evolution.

As indicated in Table 1, there are at least nine other smaller satellites with estimated diameters ranging from about 30 to 220 km. Voyager 1 determined approximate shapes for only 1980 S28, which is just outside the main rings, and the two satellites, 1980 S1 and 1980 S3, which share essentially a common orbit. All three were found to be distinctly non-spherical. Voyager 2 will determine the shape of all the remaining known satellites, including the F-ring shepherding satellites (1980 S26 and 1980 S27), the two satellites recently discovered in ground-based images[5–7] which share a common orbit with Tethys (1980 S13 and 1980 S25) and 1980 S6, which shares an orbit with Dione. There will also be a search for new satellites, both outside and embedded within the ring system. Photometrically accurate measurements of the amount and degree of polarization of light scattered from the different satellites will provide additional information on the texture and gross composition of their surfaces.

Plasma

The fourth major area of study is that of the planetary magnetic field, plasma, trapped energetic particles and wave emissions comprising Saturn's magnetosphere. Periodic bursts of radio emissions in the 100–500-KHz frequency range provided the first accurate determination of the period of rotation of Saturn's magnetic field, which is believed to be the period of rotation of Saturn's interior regions where the magnetic field originates. This period, which is 10 h 39.4 min, will be more accurately determined with the longer time-base provided by Voyager 2, while measurements of Saturn's magnetic field may identify the longitudinal asymmetry resulting in the localization of the radio emission to specific saturnian longitudes. Voyager 2 will also provide additional observations of the 2.7-day modulation of the radio emissions thought possibly to result from Dione's interaction with the magnetosphere.

The effects of Mimas and Enceladus on the magnetosphere will be directly measured as Voyager 2 passes just inside Mimas's orbit. Measurements deep within the magnetosphere will better define the structure of the equatorial disk of plasma and the extent to which the icy satellites are sources or absorbers of magnetospheric particles. From Voyager 1, it is known that Titan's atmosphere is the source of at least some of the particles.

Voyager 2 may also provide a rare opportunity for observing the effects of Jupiter's magnetotail on Saturn's magnetosphere. There is recent evidence from Voyager 2 that Jupiter's magnetotail, which may resemble a tattered wind sock flapping in the solar wind, extends at least 4.5×10^8 km outwards from Jupiter. Because, during the past few months, Saturn has been almost radially aligned with Jupiter, it is likely that Saturn has been occasionally immersed in the extended jovian magnetotail. Such an immersion could at least temporarily result in variations in the size, shape and other characteristics of Saturn's magnetosphere, leading to the possibility that there are significant differences between the Voyager 2 observations and those of Pioneer 11 and Voyager 1.

Outcome

Intensive observations of the saturnian system, which began on 5 June, will continue until 28 September, when Voyager 2 will return to the cruise mode and, like Voyager 1, continue exploring the interplanetary medium at increasing distances from the Sun. In late 1985, Voyager 2 will return to the encounter mode as it approaches a flyby of Uranus on 24 January 1986, and heads towards an encounter with Neptune on 24 August 1989. Beyond Neptune, both spacecraft will continue exploring the outer limits of the region dominated by the solar wind, where low-energy galactic cosmic rays may be first observed and beyond which lies the interstellar medium. Thus, there is potential for a continued and significant scientific return from the Voyager spacecraft well beyond that originally planned for the mission to Jupiter and Saturn.

1. *Science* **207**, 400–453 (1980).
2. *J. geophys. Res.* **85**, 5651–5959 (1980).
3. *Science* **212**, 159–243 (1981).
4. *Nature* **292**, 675–755 (1981).
5. Seidelmann, P. K. *et al.* Preprint (1981).
6. Veillet, C. *IAU Circ.* No. 3593, 3608 (1981).
7. Larson, S. M. *Icarus* (in the press).

© 1981 Nature Publishing Group

Thermal structure and dynamics of Saturn and Jupiter

J. A. Pirraglia*, B. J. Conrath*, M. D. Allison† & P. J. Gierasch‡

* NASA/Goddard Space Flight Center, Code 693.2, Greenbelt, Maryland 20771, USA
† Department of Space Physics and Astronomy, Rice University, Houston, Texas 7701, USA
‡ Laboratory for Planetary Studies, Cornell University, Ithaca, New York 14853, USA

High resolution Voyager IRIS measurements for Saturn and Jupiter are assembled in meridional cross-sections of the retrieved upper tropospheric temperatures. The calculated thermal wind shear in the upper troposphere is highly correlated on both planets with the cloud top winds derived from imaging data. In contrast, temperatures below ~300 mbar are not simply related to the zonal jet structure. The upper tropospheric temperatures seem to have been more consistently correlated with cloud top winds than with major albedo features at the time of the Voyager encounters.

THE atmospheres of Saturn and Jupiter are the most energetic observed in the Solar System. Both planets are encircled by high velocity super-rotating equatorial currents and an attendant system of axially-symmetric westward and eastward jet streams at higher latitudes. While differing in their axial tilt, gravitational potential, emission temperature and visual contrast, Saturn and Jupiter are approximately similar in size, rotational period, and gross composition. Despite these similarities major differences in their dynamical structures are observed. Smith *et al.*[1] report that the equatorial prograde velocities measured relative to the radio period are much larger and wider in latitudinal extent on Saturn than on Jupiter. Also, with one exception, anticyclonic shear on Jupiter is generally well correlated with brighter cloud bands or zones while on Saturn no such correlation is evident.

A preliminary analysis[2] of the saturnian atmosphere by the Voyager IR instrument (IRIS)[3] has revealed substantial latitudinal structure in upper tropospheric temperatures. This structure is further investigated here by constructing a meridional temperature cross-section and, for comparison, similar cross-sections were obtained using IRIS data for Jupiter. Correlation of the structure of the upper troposphere with the cloud level dynamics is explored for both planets by comparing the computed thermal wind shear with the imaging measurements of zonal wind velocites.

Zonal thermal cross-sections

A meridional thermal cross-section for Saturn's northern hemisphere is shown in Fig. 1*a*: temperature profiles retrieved between 730 and 150 mbar from 82 individual spectra taken with roughly uniform distribution in latitude but at various longitudes were assembled. For these measurements the IRIS field of view projected onto the planet varied between 2° and 5° in latitude. The vertical resolution for the retrieved temperature is ~±0.5 scale heights and the error in temperature due to instrumental noise for each spectral measurement is ~±0.5K. Continuous high resolution latitudinal coverage was available only in the northern hemisphere.

Thermal cross-sections between 140 and 500 mbar for Jupiter's northern and southern hemispheres, assembled from 150 and 166 Voyager 1 spectral measurements respectively, are shown in Fig. 1 *b*, *c*. The field of view projected onto the planet, vertical resolution, and thermal noise level for these data are comparable with those for Saturn.

The pointing data for these measurements were provided by the Voyager supplementary experimenter data record (SEDR)[4]. In some cases it has been possible to check this pointing information with reference to simultaneously shuttered Voyager imaging records. In these cases agreement between the imaging and the SEDR has been within 1° latitude. Although such verification is impossible for spectral measurements taken on the dark sides of the planets, the estimated residual 3σ pointing error is expected to be within 1°–3° of latitude, depending on the projection of the field of view onto the planets. Latitudinal correlations between the thermal and visual observations may therefore be sought within these error bounds.

The gross structures exhibited in the cross-sections for both Saturn and Jupiter show horizontal scales of variation comparable with the width of the zonal jet streams at mid-to-high-latitudes. The direct visual comparison of latitudinal thermal variation in these cross-sections with both albedo and wind velocity is more difficult. The Saturn cross-section (Fig. 1*a*) shows no simple correlation with the visual albedo variation measured by Voyager imaging[1], but some correlation is observed between the visual albedo and the structure of Fig. 1 *b*, *c* for Jupiter at upper levels near 150 mbar. Higher albedo is in most places associated with lower temperatures.

The upper tropospheric structure for Jupiter seems to be in rough agreement with the IR measurements of Pioneer 11 reported by Orton *et al.*[5]. A region of special interest is that

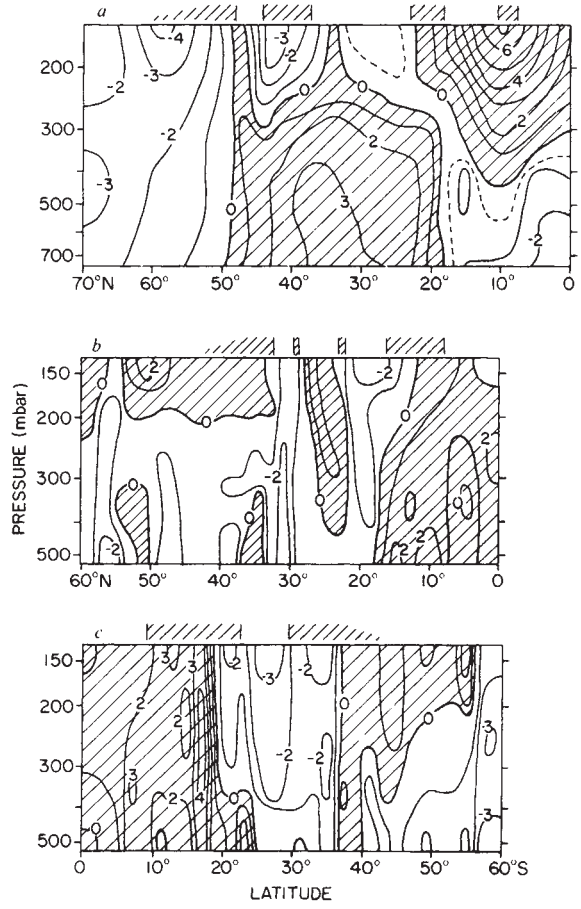


Fig. 1 Zonal cross section of deviations from the latitudinal mean retrieved temperature (K) in: *a*, Saturn's northern hemisphere as measured by Voyager 1 IRIS, November 1980. *b*, Jupiter's northern hemisphere, measured in March 1979. *c*, Jupiter's southern hemisphere, measured in March 1979. The latitudes of major albedo features are indicated schematically.

0028-0836/81/340677--03$01.00

© 1981 Macmillan Journals Ltd

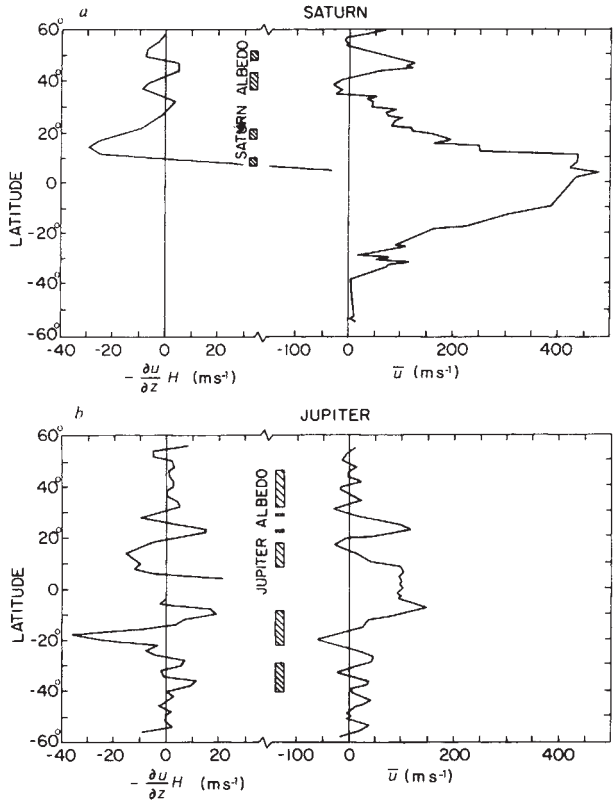


Fig. 2 *a*, Comparison between the thermal wind shear at 150 mbar and the cloud top zonal wind velocities on Saturn. *b*, Comparison between the thermal wind shear at 140 mbar and the cloud top zonal wind velocities on Jupiter obtained from Voyager 1 IRIS measurements. For convenience the negative of the wind shear multiplied by one atmospheric scale height is displayed. The major albedo contrast is also indicated.

between 24° and 30° N (historically the North Temperate Belt or NTB). According to Peek[6] this region is one of the most variable on the planet. The region was dark during the Pioneer 10 and 11 encounters and the 20μm (upper tropospheric) temperatures were correspondingly high. Figure 1*b* shows that this region was still warm at 150 mbar during the Voyager 1 encounter; it was also warm during the Voyager 2 observations. The appearance of this region during the Voyager encounters, however, showed a high albedo typical of a zone[1]. According to the Earth-based observations by R. F. Beebe (personal communication) the belt-like region began to cloud over in 1977–78 so that the zone-like appearance had been present for <2 yrs before the Voyager encounters. This interval is, however, well within the radiative response time of 10^8 s for this level of Jupiter's atmosphere[7], and the upper tropospheric temperatures may not yet have responded to the reduction of IR flux at the cloud level.

The thermal cross-sections for both planets show substantial differences in the horizontal structure between levels above and below 300 mbar. These differences may reflect inaccuracies in the temperature retrievals at low levels due to the opacity effects of latitudinally variable aerosol concentrations. If the retrieved temperatures are correct at all levels, however, then the presence of some as yet unknown driving mechanism is indicated. Possible candidates include aerosol heating in the upper troposphere, radiative screening effects associated with clouds, adiabatic heating and cooling due to vertical motion[8], and eddy or wave generated fluxes of heat and momentum.

Thermal wind shears

The observed latitudinal gradients in temperature on Saturn and Jupiter require a dynamical response, and the velocity and length scales inferred from the Voyager imaging measurements are consistent with geostrophic balance. Therefore, with the additional assumption of hydrostatic equilibrium, the vertical shear of the eastward zonal flow u is given by the thermal wind equation

$$\frac{\partial u}{\partial z} = -\frac{R}{2\Omega a H \sin\lambda}\frac{\partial T}{\partial \lambda}$$

where the derivative of temperature with respect to latitude λ is taken along an isobaric surface, and R is the gas constant. The planetary rotational frequency is $\Omega = 1.64\times10^{-4}\ s^{-1}$ and $1.76\times10^{-4}\ s^{-1}$ for Saturn and Jupiter respectively. The respective planetary radii are given by $a = 6.03\times10^4$ km and 7.14×10^4 km. The scale height H is ~38 km for Saturn and 22 km for Jupiter at upper tropospheric levels. Because of the limited spatial resolution of the IR measurements and the uncertainty in the ammonia condensation level, it is impossible to match reliably an integrated wind shear to the zonal velocities at the cloud deck. Nevertheless, an estimate of the thermal wind shear in the upper troposphere for the observed zonal temperatures on the two planets is of diagnostic value in determining the extent of the coupling between the upper level dynamics and the observed zonal winds.

Figure 2 shows the thermal wind shear derived from the temperature retrievals at 150 mbar for Saturn and 140 mbar for Jupiter. These computations used the same data sets as those used for the cross-sections in Fig. 1. For Saturn the meridional derivative was computed for every 2.5° latitude as a finite difference over 5° between temperatures given by a smooth fit to the data. For Jupiter the finite difference interval was 4° at every 2° of latitude. The imaging measurements of the zonal velocity $\bar{u}$ at the cloud tops are plotted to the right of the wind shears for comparison. The cloud top zonal winds for Saturn were taken from ref. 1 while those for Jupiter were adapted from ref. 9. The approximate latitudinal positions of the visual albedo variations at the cloud deck are also indicated in Fig. 2 with dark markings for 'belts', interspersed with the lighter 'zones'.

Wherever the sign of $-(\partial u/\partial z)H$ is the same as that for $\bar{u}$ a reduction of the cloud top velocity with height is implied. Different scales have been used for $-(\partial u/\partial z)H$ and $\bar{u}$ to aid their visual comparison. Because of the finite field of view for the measurements computed shear magnitudes are best regarded as lower limits for gradients of thermal fields which may vary significantly over scales too small to be completely resolved.

A comparison of the results for the two planets implies that the dynamical length scale, as indicated by the latitudinal change in the sign of the shear, is slightly larger for Saturn than for Jupiter. In general, there is a good correlation between the positions of most of the observed jets and the wind shear extrema. The shear is in the sense required to reduce the strength of the zonal velocities at cloud level with height, but the magnitude of the shear is such that several scale heights would be required for a significant modification of wind speeds comparable with those observed at the cloud tops.

On Jupiter, alignment between the peaks of the jets and extrema of the thermal wind shear is in most places within the ±1–3° of residual pointing uncertainty. The agreement seems to be especially good at the strong prograde jet at 23° N and throughout the alternating system of jets between 15 and 40° S. Although it is not certain how close to the equator the computed thermal wind shears are valid, they have been plotted within ±4° of the equator and some reduction in the strength of the equatorial jet with height is indicated.

The thermal wind shear at 18° S on Jupiter seems to be stronger than at any other mid-to high-latitude location. The retrograde jet at 20° S as measured by the Voyager imaging is unusually broad. The apparent enhancement in the wind shear at this location may therefore simply reflect the better spatial resolution in the thermal measurement relative to the jet width.

Note that the region between 24 and 30° N on Jupiter retains a strong correlation between the thermal wind shear and the wind field at the cloud tops. This is a consequence of the relatively warm upper troposphere even through this normally belt-like region had clouded over before the Voyager 1 encounter. Thus, the upper tropospheric temperatures seem to have been more

© 1981 Nature Publishing Group

strongly correlated with the cloud level jets than with the large scale albedo features at the time of the Voyager observations.

The computed thermal wind shear for Saturn is closely correlated in position with the retrograde jet centred near 38° N and with the prograde jet near 46° N, again in the sense required to reducethe cloud top velocities with height. The magnitude of the thermal wind shear relative to the measured $\bar{u}$, however, is much larger for the retrograde jet. There is also some correlation between the vertical shear and the weak retrograde jet centred near 55° and at least a correlation in sign for the prograde cloud motions around 32° N. The thermal wind shear between 5° and 10° N indicates some reduction in the strength of the very rapid equatorial jet with height. Between 10 and 30° N the computed thermal wind shear shows a significant departure from the general correlation observed elsewhere. Here the shear is in the sense consistent with a retrograde velocity, but the imaging measurements show this region to be a part of the prograde equatorial jet.

Smith *et al.*[1] indicate that the Saturn jets at latitudes above 35° show some correlation with the albedo variation. Unlike Jupiter, however, the peaks of the jets seem to coincide with the middle of the bright and dark bands. It is interesting that the extrema of the computed thermal wind shear also show some correlation with the middle of the bright and dark bands, not only at the same high latitudes, but also at 32° N and especially at 15° N where the strong retrograde shear corresponds closely to the middle of a bright region to the south of Saturn's north equatorial belt. Comparison of thermal wind shears with zonal velocities in Saturn's southern hemisphere must await the more comprehensive Voyager 2 IR measurements.

Conclusions

The thermal wind shear calculated from the IRIS derived temperatures in the upper troposphere is highly correlated with the imaging derived jets at most latitudes on both Jupiter and Saturn and is in the sense required for the wind to decrease with height. An exception is the strong retrograde feature inferred from the thermal data at 15° N on Saturn which does not appear in the imaging data. Contrary to expectation, the thermal structure in the lower troposphere below ~300 mbar is not well correlated with cloud top winds on either planet. This may be the result of a haze or cloud effect not taken into account in the thermal retrievals; however , the possibility that the thermal structure is real cannot be ruled out until particulate contributions to the IR opacity of the lower tropospheres of both planets is better understood.

The thermal wind shear and major albedo features have similar latitudinal scales, but the exact nature of this correlation differs on the two planets. On Saturn, the extrema in the wind shear tend to fall near the centres of the bands while on Jupiter they occur near the interfaces of belts and zones. The observation of the anomalously bright but warm region between 24° and 30° N on Jupiter suggests a better correlation between upper tropospheric temperatures and the jet streams than with the large scale albedo features at the time of the Voyager encounters.

We thank Rudolf Hanel, principal investigator for the Voyager IRIS team, for valuable comments and suggestions.

Received 22 May; accepted 20 July 1981.

1. Smith, B. *et al. Science* **212**, 163–191 (1981).
2. Hanel, R. *et al. Science* **212**, 192–200 (1981).
3. Hanel, R. *et al. Appl. Opt.* **19**, 1391–1400 (1980).
4. Hanel, R. *et al. Voyager Infrared Interferometer Spectrometer and Radiometer – Documentation for Reduced Data Records for Jupiter* Document X-693-821-8 (NASA Goddard Space Flight Center Greenbelt, 1980).
5. Orton, G. S. *et al. Icarus* (submitted).
6. Peek, B. M. *The Planet Jupiter* (Faber, London, 1958).
7. Gierasch, P. J. & Goody, R. M. *J. atmos. Sci.* **26**, 979–980 (1969).
8. Conrath, B. J. *et al. J. geophys. Res.* (in the press).
9. Ingersoll, A. P. *et al. J. geophys. Res.* (in the press).

Morphology of Saturn's aurora

B. R. Sandel & A. L. Broadfoot

Earth and Space Sciences Institute, University of Southern California, Tucson Laboratories, Tucson, Arizona 85713, USA

Aurorally-excited emissions of atomic and molecular hydrogen come from a narrow circumpolar band near 80° north and south latitude on Saturn. The aurorae, which lie near the edge of the polar cap region, are continuously excited in both the north and south. If the strong variations observed in the auroral intensity are temporal, rather than longitudinal, they may be related to the periodic structure in the Saturn kilometric radiation.

BECAUSE Saturn possesses a dense atmosphere and a strong magnetic field, it is reasonable to expect that aurorae should also be present. Before the encounter of Voyager 1, two lines of evidence suggested the presence of polar aurora on Saturn, but both admitted alternate interpretations. Pioneer 11's ultraviolet photometer showed enhanced emission in its long wavelength channel when the field of view crossed the polar limbs, as compared with the centre of the disk[1]; this could have been due to auroral excitation or to limb brightening. Spatially-resolved images of Saturn obtained by the International Ultraviolet Explorer (IUE) satellite showed enhancements in HLyα emission which could have come from Saturn's polar atmosphere[2]. At different times the enhancements appeared at both north and south poles, but were sometimes completely absent. The polar enhancement ranged from 300 to 900 R averaged over the resolution element of 6×10 arc s. The source of the emission could not be localized to the atmosphere of Saturn, and it was noted that the observations were consistent with scattering of sunlight by a sporadic cloud of neutral hydrogen associated with Saturn's rings. The first unambiguous evidence[3] for saturnian aurora was provided by the UV spectrometer (UVS) on board Voyager 1. Aurorally excited emissions were found at HLyα and in the Lyman and Werner bands of H_2 near Saturn's poles. We now expand that brief description[3] of the aurora.

The UVS is an objective grating spectrometer covering the wavelength range 530–1,700 Å in 126 contiguous channels[4,5]. This wavelength region includes the atomic HLyα line that can be excited by resonance scattering or particle impact, and the particle-excited Lyman and Werner bands of molecular hydrogen. During the encounter the UVS was used to map the planet at varying spatial resolutions.

Of the data examined so far, the polar scan sequence shown in Fig. 1 has yielded the best information on the position of the aurora. The path of the slit intersected an auroral oval twice, once on each side of the pole. The presence of the aurora was defined by locally enhanced emissions at HLyα and in the H_2 bands. In the polar cap region, the auroral emission was reduced by a factor of at least 4, and may have been entirely absent. To determine the width and location of the auroral excitation, the data from the crossing on the near side of the pole have been compared with a simple model of the distribution of auroral emission. This model shows that the auroral emission extends

0028–0836/81/340679—04$01.00

© 1981 Macmillan Journals Ltd

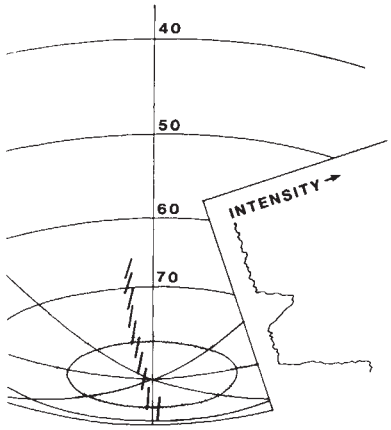


Fig. 1 Geometry of a polar auroral scan. The range is 3.6 R_S and the latitude of the spacecraft is −36.6°. Dashes show the position of the 0.1° × 0.86° UVS slit at 1.5-min intervals as the slit scanned across the south pole and onto the dark limb. The stippled area shows the inferred location of the source of auroral emission between 78° S and 81.5° S. The slit crossed the aurora on the near side of the pole at λ_{SLS} = 190° near the dusk terminator. H_2 band emission along the path of the slit is shown in the inset.

between 78 and 81.5° S. On the far side of the pole, the emission occurs very close to the limb of the planet, and is difficult to locate accurately. Furthermore, on the far side, the uncertainty in latitude introduced by parallax and the uncertain altitude of the auroral emission is much greater. No information on the distribution of the intensity within the aurora is yet available. The average apparent brightnes is ~5 kR based on the intensity of the 1,105-Å H_2 band feature, which should be optically thin[5]. Because the aurora was observed at an angle of 56° from the local zenith, the nadir intensity was ~2.8 kR. Figure 1 shows the auroral oval at constant latitude; this is consistent with, but not required by, the data. We have also anticipated a result derived later—that the aurora is at least roughly continuous in longitude, to extend the auroral oval around the planet in longitude.

In the north, the aurora occurs at nearly the same location, as the nearly dipolar, axisymmetric nature[6] of the magnetic field would imply. Approximately 11 h after closest approach the north polar area was searched for aurora. The geometry was similar to that shown in Fig. 2b, except that the spatial resolution was a factor of 3 better. Two scans near local midnight show that the boreal aurora lies poleward of 76° N, and may not extend below 78° N. These scans do not define the poleward extent of the excitation. We assume that the boreal and austral aurora have the same latitudinal extent, as their equatorward boundaries fall at nearly the same latitude, and the magnetic field has north–south symmetry. The two scans show a 50% increase in H_2 band emission (from 2.5 to 3.7 kR) from the first to the second, 20 min later. As the planet rotated only 13° in this time, and because the field of view was positioned near the 'elongation' of the auroral oval, the longitude range within the slit did not change significantly between scans. The change in measured intensity is probably due to a temporal variation in auroral brightness.

More extensive observations at lower spatial resolution are available for study. The north–south map (NSM) sequence was executed ~36 h before and after closest approach. This observing sequence, shown in Fig. 2, consisted of a series of swaths from north pole to south pole, approximately along the central meridian of the planet. Each swath of the pre-encounter NSM was made up of 12 steps in latitude, each step lasting 48 s. Swaths were repeated at intervals of 12 min, or 7° in rotation phase, during an almost complete rotation period of the planet. Saturn was therefore mapped in latitude (by stepping the scan platform) and in longitude (by taking advantage of the planet's rotation).

The NSM sequences show that aurorae are present near both poles, and that they are probably continuous, at least during the few days nearest encounter. Because the apparent size of the auroral oval is small, the fact that the slit does not always scan exactly along the central meridian is important. The field of view was well centred on the central meridian at the beginning of the post-encounter NSM, but drifted partially off the aurora to the east during the 10-h sequence. Therefore we cannot derive strong conclusions about time or longitudinal variations over the full span of the observation, but the data clearly demonstrate the presence of aurora near the north pole throughout the NSM sequence. The favourable configuration shown in Fig. 2*b* was maintained for the first 3 h of the sequence. During this time the brightness of the H_2 bands decreased from its initial level of 20 kR by a factor of ~2. This change can be interpreted as a temporal variation or as a longitudinal variation in the brightness of the aurora. The field of view sampled the auroral oval on both sides of the pole, so the presence of a brighter area near both λ_{SLS} = 25° and 200° is compatible with the data. (The longitude system used here, λ_{SLS}, is based on a sidereal period of 10 h 39.4 min defined by the Voyager planetary radio astronomy experiment[7].)

A similar well-defined variation was recorded during the pre-encounter NSM. Part of the data for which we can rule out significant changes in signal level induced by changes in the position of the field of view are shown in Fig. 3. The H_2 band intensity increased by a factor of 5, remained high for 2 h (or ~70° in Saturn rotation phase) then returned to its lower level. At earlier and later times, the signal was about the same as the lower level shown in Fig. 3, but at these times the signal could have been modulated to some extent by changes in the position of the field of view. Note that the longitude of the field of view was ~160° at the peak in H_2 band emission. This differs from the position quoted earlier[3] because we have now included the fact

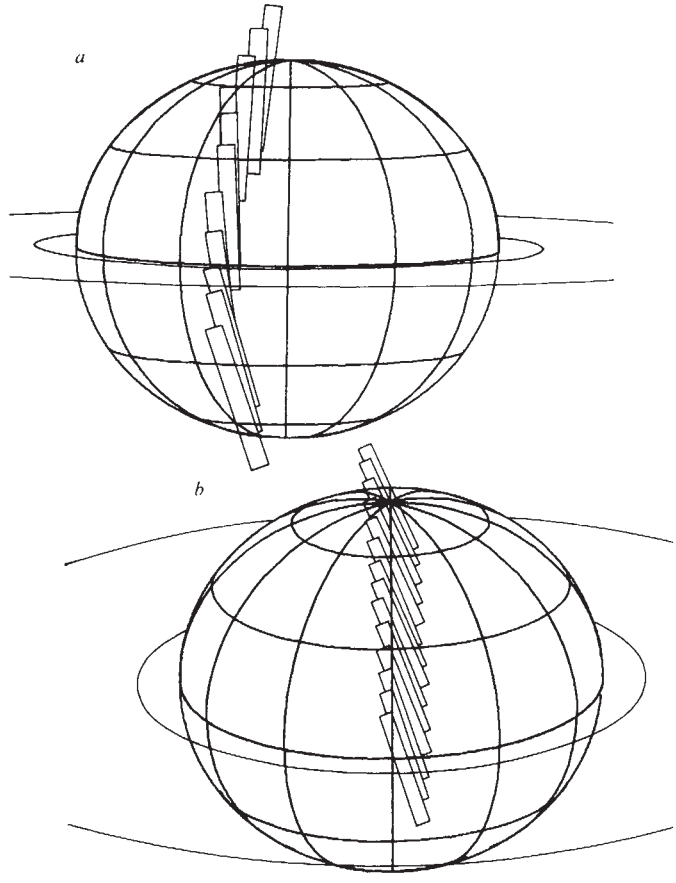


Fig. 2 Views from the spacecraft showing the geometry of the North–South Map sequence. The pre-encounter NSM *a* was executed when the latitude of the spacecraft was 4.8° N. Both boreal and austral aurorae were visible, although the austral aurora was quite near the limb of the planet. Near the poles, the slit was near 10.00–11.00 LT. For the post-encounter NSM *b*, the latitude of the spacecraft was 22° N and the latitude of the austral aurora was not included in the scan. The rectangles show the positions of the UVS slit during a typical swath. Saturn's central meridian is near 3.00 LT.

© 1981 Nature Publishing Group

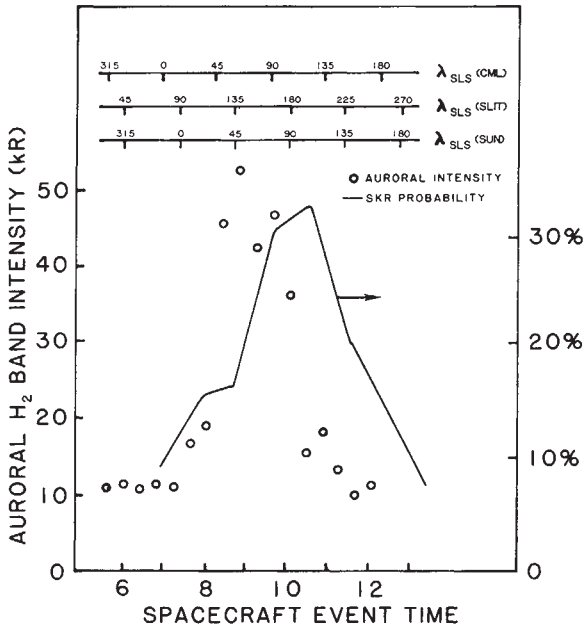


Fig. 3 H_2 band intensity as a function of time during a portion of the pre-encounter NSM. The intensity integrated over the Lyman and Werner bands of H_2 has been determined from the measured intensity of the 1,105 Å feature, which is relatively unaffected by self-absorption. The solid line shows the probability for detection of Saturn kilometric-wave radiation[8].

that the field of view is not on the central meridian. At the same time, the austral aurora showed a qualitatively similar variation. The unfavourable observing geometry (see Fig. 2*b*) precludes a quantitative analysis of the austral auroral intensity based on the NSM data. Figure 3 also shows the probability of occurrence of the Saturn kilometric radiation (SKR) taken from Gurnett *et al.*[8]. Following the discovery of the SKR by Kaiser *et al.*[9], the probability of observing this radiation was found to be strongly correlated with the longitude of the Sun in the λ_{SLS} system[8,10]. For the comparison in Fig. 3, the longitude of the Sun was calculated as a function of spacecraft time, and then the relationship between SKR probability and solar longitude was used to plot the SKR probability on the same time scale as the auroral brightness. In spite of a phase difference of ~50°, the correlation between the SKR probability and the auroral intensity apparent in Fig. 3 is good enough to suggest that the two phenomena may be closely linked. A test of this correlation must await the Voyager 2 encounter.

Spectra of the aurora are compared in Fig. 4. The signature of the H_2 Lyman and Werner bands is present between ~900 Å and 1,130 Å. The latitude of the spacecraft was different when the two spectra of Saturn's aurora were taken, implying different pathlengths through the atmosphere to the auroral source. Stronger self-absorption by H_2 in the longer atmospheric pathlength of the middle spectrum probably accounts for the gross differences shortward of 1,050 Å in the Saturn spectra (see Fig. 9 of ref. 5). The differences between the auroral spectra of Saturn and Jupiter between 1,100 and 1,200 Å parallel those in the dayglow spectra of the two planets[3].

The power required to drive the aurora was estimated earlier[3] to be $\sim 2\times10^{11}$ W. This was based on an average brightness of 5 kR in the H_2 bands extending over an area of 4.6×10^{18} cm^2 (the surface area of Saturn between 78° and 81.5° latitude) and the efficiency for electron excitation of the H_2 bands[5]. This re-analysis shows no basis for altering the early estimate, but it does emphasize that the presence of temporal and/or longitudinal variations in the surface brightness of the aurora may bias the apparent average brightness. About two thirds of the auroral information available from the Voyager 1 encounter has been extracted.

Other sources of uncertainty are important. The brightness over all wavelengths in the H_2 Lyman and Werner band systems has been estimated using the measured intensity of the 1,105 Å feature. This feature is relatively unaffected by self-absorption in long slant-paths through the atmosphere to the auroral source and hence more likely to represent the true emission intensity. The uncertainty in the model relating the total H_2 band emission to the 1,105 Å feature may be as much as a factor of 2 (D. E. Shemansky, personal communication). Further uncertainty is introduced when the apparent brightness of an aurora near the limb of the planet must be used to calculate the surface brightness. Hence we have given more weight to the intensity measurements made at the smallest atmospheric pathlengths in assigning the average auroral brightness of 5 kR. Because of these uncertainties, figures such as auroral power dissipation are uncertain by a factor of at least 4. On the other hand, the relative brightness measured within a given sequence (for which the observing geometry remains constant) depends mainly on the pointing accuracy and should be very good. The location of the aurora, at least at the time of the observation in Fig. 1, is well defined to within at most 2° in latitude. Finally, the emissions in the H_2 bands and at HLyα from a thin layer above the sub-solar limb of Saturn[3] are too weak to contaminate these auroral measurements significantly.

The aurora described here are probably related to the bursts of HLyα emission reported by Clarke *et al.*[2]. The HLyα component of the auroral emissions is generally a factor of ~5 fainter than the H_2 band emission. The brightest HLyα emission measured in the UVS data was 9 kR observed during the pre-encounter NSM. Assuming that the radiation reported by Clarke *et al.* arose from either the northern or the southern aurora defined by the UVS, we estimate that the true surface brightness corresponding to the bursts reported by Clarke *et al.* ranges between 13 and 40 kR. The brightest levels were observed in only one of the seven observing periods distributed over 6 months; their lowest level bursts detected correspond roughly in intensity to the brightest HLyα emission measured to date by Voyager.

The morphology of Saturn's aurorae resembles that of Earth's aurorae more than that of Jupiter's aurorae. The edge of Saturn's polar cap region lies at a latitude of 78.7°–75° (ref. 6), and thus corresponds closely to the location of the auroral excitation, as in the case of the Earth. This suggests that analogous mechanisms supply energy for the aurora on the two

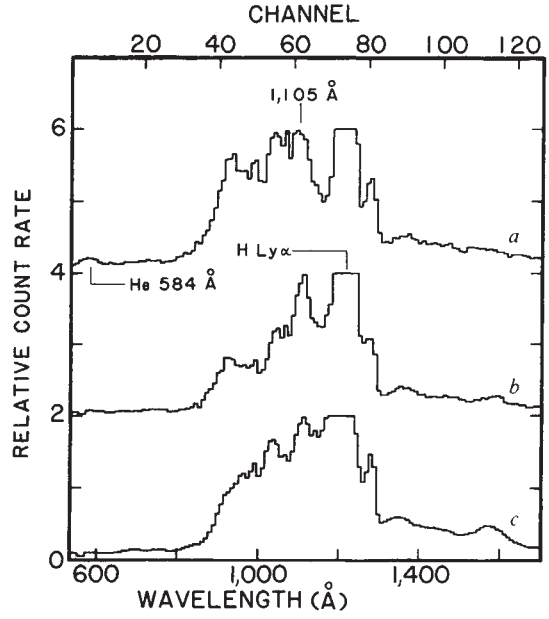


Fig. 4 Auroral spectra. *a*, Saturn aurora, observed at a local zenith angle 60°. *b*, Saturn aurora, local zenith angle 76°. *c*, Jupiter aurora. Aurorally-excited emissions in the Lyman and Werner bands of H_2 are apparent between 900 and 1,130 Å. The spectra have been normalized to equal levels at 1,105 Å. The Saturn spectra differ markedly between 850 Å and 1,050 Å because of different absorption pathlengths. The 1,105 Å feature marked was used as an indicator of the intensity radiated in the H_2 bands. The HLyα line has been truncated.

© 1981 Nature Publishing Group

planets. On the other hand, the equatorward boundary of Jupiter's aurorae matches the magnetic projection of Io's plasma torus onto the atmosphere, suggesting excitation by an interaction between the atmosphere and the plasma torus[11]. (Aurorae may be present nearer the poles, at the boundaries of the polar caps, as well.) The energy deposited in the Earth's atmosphere by aurorae is correlated with geomagnetic activity, ranging over more than 2 orders of magnitude[12]. The average value[13] is probably near 1.7 erg cm^{-2} s^{-1}, which corresponds to about 3×10^{10} W, roughly 1/7 of the power deposited at Saturn. The interaction between the magnetosphere of a planet and the solar wind, the ultimate source of auroral energy on Earth and probably on Saturn, is important in determining the power available to auroral processes. This interaction transfers about 10^{11} W into Earth's magnetosphere[14], about 1/10 of the 10^{12} W injected into the saturnian magnetosphere, calculated using measurements of the saturnian magnetic field[6] and general scaling relationships for magnetospheric parameters[14]. Thus the ratio of power in aurora on the two planets is about the same as the ratio of the power transferred to the magnetosphere from the solar wind. This, and the occurrence of aurora near the edges of the polar caps on both planets, suggest that the aurorae result from similar magnetospheric processes.

This work was supported by the JPL, California Institute of Technology under NASA contract NAS 7-100. Additional support was provided by the Planetary Sciences Discipline of NASA's Office of Space Sciences.

Received 13 May; accepted 8 July 1981.

1. Judge, D. L., Wu, F.-M. & Carlson, R. W. *Science* **207**, 431 (1980).
2. Clarke, J. T., Moos, H. W., Atreya, S. K. & Lane, A. L. *Nature* **290**, 226 (1981).
3. Broadfoot, A. L. *et al. Science* **212**, 206 (1981).
4. Broadfoot, A. L. *et al. Space Sci. Rev.* **21**, 183 (1977).
5. Broadfoot, A. L. *et al. J. geophys. Res.* (in the press).
6. Ness, N. F. *et al. Science* **212**, 211 (1981).
7. Desch, M. D. & Kaiser, M. L. *Geophys. Res. Lett.* (in the press).
8. Gurnett, D. A., Kurth, W. S. & Scarf, F. L. *Science* **212**, 239 (1981).
9. Kaiser, M. L., Desch, M. D., Warwick, J. W. & Pearce, J. B. *Science* **209**, 1238 (1980).
10. Warwick, J. W. *et al. Science* **212**, 239 (1981).
11. Sandel, B. R. *et al. Science* **206**, 962 (1979).
12. Murphree, J. S. & Anger, C. D. *Geophys. Res. Lett.* **5**, 551 (1978).
13. Torr, D. G., Torr, M. R., Hoffman, R. A. & Walker, J. G. C. *Geophys Res. Lett.* **3**, 305 (1976).
14. Siscoe, G. L. *Icarus* **24**, 311 (1975).

Saturn ionosphere: theoretical interpretation

S. K. Atreya* & J. H. Waite Jr†

* Department of Atmospheric and Oceanic Science, Space Physics Research Laboratory, The University of Michigan, Ann Arbor, Michigan 48109, USA; † Marshall Space Flight Center, Huntsville, Alabama 35812, USA

Voyager 1 high latitude and Pioneer 11 equatorial ionospheric structure indicate a solar EUV-controlled ionosphere with a possible molecular ion in the topside. Vibrationally excited H_2 in the high latitudes may be an important loss mechanism. Dynamical effects are expected to be important for determining the peak density and its location.

THE Voyager 1 coherent dual-frequency radio system (RSS) at wavelengths 3.6 and 13 cm provided the first look at a high latitude (~73 °S) ionospheric profile on Saturn[1]. The entry data reveal that near the evening terminator (solar zenith angle, $\xi = 89°$), the peak electron density of 2.3×10^4 electrons cm^{-3} occurred at a height of ~2,500 km from the 1 bar pressure level. Earlier observations on Pioneer Saturn measured a peak electron density of 1.1×10^4 cm^{-3} at 1,800 km in the 10 °S latitude range[2]. The observed ionospheric characteristics point to a markedly different ionosphere from that expected on conventional theoretical grounds.

The RSS ionospheric measurements were accompanied by a simultaneous measurement of the neutral atmosphere by means of the Voyager UV spectrometer (UVS). Modelling studies suggest that the observed ionosphere is controlled principally by the ionization of the neutrals caused by the solar EUV.

The UVS solar occultation measurements yield a neutral exospheric temperature of 820 ± 100 K at 62,000 km (ref. 3). The large size of the Sun (~500 km) projected on Saturn means that the height level at which the temperature reaches its exospheric value cannot be determined accurately. The exospheric temperature, however, is nearly the same as the topside plasma temperature (750 K)[1].

Preliminary analysis of the UVS stellar occultation data indicates that the temperature is essentially isothermal (~150–200 K) in the mesosphere up to ~750 km ($H_2 \simeq 5\times10^{13}$ cm^{-3}) and an average temperature gradient of ~0.55 K km^{-1} prevails above this altitude resulting in an exospheric temperature of ~825 K at 2,000 km above the 1-bar pressure level. Indeed, the Saturn stellar occultation transmission characteristics are similar to those of the Jupiter stellar occultation data collected on Voyager 2 (refs 4, 5) when the height scale is corrected for the lower acceleration due to gravity on Saturn. With this information and the hydrostatic law, a working model of the atmospheric density profile at the ionospheric heights can be constructed. An important caveat is that the Voyager ionosphere was measured at high latitudes, while the neutral atmosphere used here is essentially from equatorial to mid-latitude measurements[3]. The auroral activity on Saturn, however, is confined to the latitude band 78°–81.5 °S which is well outside the latitude region (73 °S) of the ionospheric measurements discussed here. Therefore, no substantial change is expected in the neutral atmosphere at the ionospheric heights from the low to the high latitudes on Saturn.

Knowledge of the atmospheric mixing as represented by the eddy diffusion coefficient, K, is important for determining the distribution of the hydrocarbons which act as an important sink to the major ions, H^+, H_2^+ and H_3^+ (refs 6, 7). K_h at the homopause may be determined from the atomic hydrogen abundance deduced from the planetary Lyα intensity[4,8,9]. The UVS data yield 3.3 kR for Saturn Lyα dayglow[3] implying 5×10^{16} cm^{-2} H atoms above the homopause, and consequently K_h of $(1.5\text{–}5.0)\times10^5$ cm^2 s^{-1}.

The hydrocarbon density distributions were calculated considering CH_4 photolysis with the eddy diffusion coefficient varying inversely as the square root of the atmospheric number density[7].

The important chemical reactions for the Saturn ionosphere were taken from previous work[6,10]. We consider here low-pressure, low plasma temperature (T_e) approximation for the radiative recombination reaction rate of the major topside ion, H^+; the new rate is 4×10^{-12} $(250/T_e)^{0.7}$ cm^3 s^{-1} (ref. 11). The calculated ionospheric profile with a comparison with the Pioneer equatorial[12] and Voyager high latitude[1] data are shown in Fig. 1. Protons (H^+) comprise the major ions in altitudes >750 km. Below 750 km, a transition to H_3^+ and heavy hydrocarbon ions such as $C_2H_5^+$ and CH_5^+ occurs. Conversion to the higher order hydrocarbon ions, $C_3H_5^+$ and $C_4H_5^+$, is also probable in the lower ionosphere[8].

Although the Pioneer and Voyager data differ in details, there is general agreement about the location and magnitude of the peak, and the extent of the ionosphere. The observed peak electron density in the Voyager data is about a factor of 10 lower than the calculated one. A similar low density jovian ionosphere profile measured by Voyager 1 was interpreted as being due to the reaction of H^+ with vibrationally excited H_2 (ref. 10). The

0028–0836/81/340682—02$01.00

© 1981 Macmillan Journals Ltd

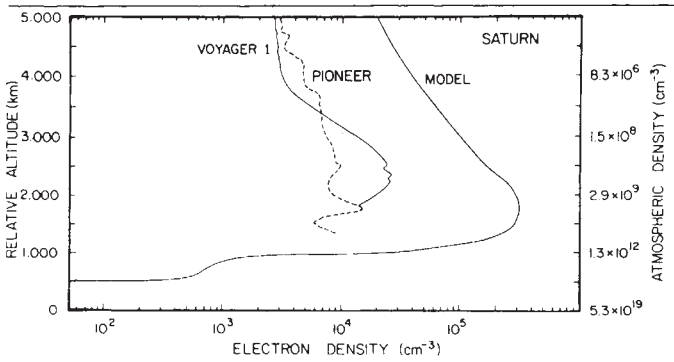


Fig. 1 Model calculations and observations of the Saturn ionosphere. The Pioneer data refer to low latitude (~10 °S), while the Voyager data are for the high latitude (~73 °S).

relatively low exospheric temperature (~800 K) on Saturn results in a small population of H_2 in $v' \geq 4$ vibrational state so that the charge exchange reaction between H^+ and H_2 ($v' \geq 4$) would not be expected to be as important on Saturn as on Jupiter. However, the H_2 vibrational levels may be non-thermally populated due to electron precipitation processes. Such non-thermal vibrational distributions were suggested by Cravens[13] from modelling of jovian electron precipitation processes. Such processes could be important in interpreting the Voyager/Saturn data provided that the effects of auroral precipitation extend beyond the observed latitude band in which significant (~5 kR) H_2-Lyman and Werner band auroral emission have been observed by the Voyager UVS[3]. The observed H_2 auroral emission imply an energy input of ~0.5 erg cm^{-2} s^{-1} (in the 78°–81.5 °N and S latitudes) on Saturn which is about a factor of 10 lower than in the jovian auroral region.

An electron precipitation event composed of 2 keV electrons with a flux of 1.5×10^8 cm^{-2} s^{-1} could reproduce the 3–5 kR of H_2-Lyman and Werner band emission seen by the UVS[3]. Such an aurora produces a model ionosphere with a peak electron density of 2×10^6 cm^{-3} near 2,400 km when the effects of the charge exchange reaction ($H^+ + H_2$ ($v' \geq 4$)) are omitted. However, if a rate constant of 10^{-9} cm^3 s^{-1} is assumed for the reaction of H^+ with H_2 ($v' \geq 4$), and a vibrational temperature of ~2,000 K, the peak electron density can be brought into rough agreement with the RSS high latitude ionosphere measurements.

Even if the loss of H^+ by charge exchange with H_2 ($v' \geq 4$) could explain low peak electron density in the high, non-auroral latitudes, similar low peak electron densities have been measured by the Pioneer 11 spacecraft near 10 °S latitude[2,12] also, and they apparently need an alternative explanation due to the unlikely occurrence of a significant vibrationally excited H_2 population in the Saturn equatorial ionosphere. The 700 R intensity of the non-auroral H_2 bands measured by the UVS[3] implies a planetwide precipitation of electrons resulting in a relatively low energy deposition rate of 1.3×10^{-2} erg cm^{-2} s^{-1} on Saturn. Relatively good agreement between the Pioneer equatorial and the Voyager high latitude Saturn ionospheric data suggests a common H^+ loss process missing from current theoretical models. Moreover, the loss of the major ion H^+ after the charge exchange with vibrationally excited H_2 ($v' \geq 4$) even in the high latitudes is only a remote possibility as the measured ionosphere is ~5°–10° removed from the auroral latitude band.

The most obvious choice for reconciling the calculations and the measurements is an ion–molecule reaction which would convert some of the H^+ into relatively short-lived molecular ions. Although methane reacts rapidly with H^+ to form molecular ions, it is not distributed high in the atmosphere due to the relatively low value of the eddy diffusion coefficient, $K_h \sim 10^6$ cm^2 s^{-1}. Even an extremely high K_h (~10^9 cm^2 s^{-1}) falls short of the required H^+ loss rate[7]. An alternative explanation is that water vapour in the rings could provide an important loss to the ionospheric H^+ (ref. 14). The OH would, however, be concentrated in the region near the rings suggesting a latitudinal variation in the $OH + H^+$ loss process, a variation not supported by the Pioneer and Voyager data. Furthermore, the large OH production rates might force unacceptable limits on the longevity of the rings. We also find that the depletion of the peak electron density due to the ring shadow occurs in low latitudes; the effect, however, is minimal (≤10% depletion in the electron density) at the latitudes of the Pioneer observations[7].

There is no adequate common explanation for the low peak electron densities observed at both the low and high Saturn latitudes. Dynamical effects may have an important role in the low latitude ionosphere and auroral effects may be important at high latitudes. However, some as yet unknown H^+ loss process may be the controlling factor at all Saturn latitudes.

The location of the peak in electron density seems to be ~500 km higher than calculated (Fig. 1); this is, however, questionable as the calculated peak is quite broad. But many theories for reconciling this apparent discrepancy are possible. Calculations indicate that an average diurnal vertical drift velocity of ~14 m s^{-1} for H^+, produced by an electrical field of ~3 mV m^{-1}, southward neutral winds of ~140 m s^{-1}, and/or vertical winds of ~14 m s^{-1} are capable of raising the peak by 500 km (refs. 7, 15).

This research was supported by a grant to the University of Michigan from the Planetary Atmospheres Program of the NASA Solar System Exploration Division.

Received 13 May; accepted 1 July 1981.

1. Tyler, G. L. *et al. Science* **212**, 201–206 (1981).
2. Kliore, A. J., Lindal, G. F., Patel, I. R., Sweetnam, D. N. & Hotz, H. B. *Science* **207**, 446–449 (1980).
3. Broadfoot, A. L. *et al. Science* **212**, 206–211 (1981).
4. Atreya, S. K., Donahue, T. M. & Festou, M. C. *Astrophys. J. Lett.* **247**, L43–L47 (1981).
5. Festou, M. C. *et al. J. geophys. Res.* **86**, 5715–5725 (1981).
6. Atreya, S. K. & Donahue, T. M. *Jupiter* (ed. Gehrels, T.) 304–318 (The University of Arizona Press, 1976).
7. Waite, J. H. Jr thesis, Univ. Michigan (1981).
8. Hunten, D. M. *J. Atmos. Sci.* **26**, 826–834 (1969).
9. Wallace, L. & Hunten, D. M. *Astrophys. J.* **182**, 1013–1031 (1973).
10. Atreya, S. K., Donahue, T. M. & Waite, J. H. Jr *Nature* **280**, 795–796 (1979).
11. Bates, D. R. & Dalgarno, A. *Electronic Recombination, Atomic and Molecular Processes* (ed. Bates, D. R.) 245 (Academic, New York, 1962).
12. Kliore, A. J. *et al. J. geophys. Res.* **85**, 5857–5870 (1980).
13. Cravens, T. D. thesis, Harvard Univ. (1974).
14. Shimizu, M. *Proc. 13th Lunar planet. Symp.* (Institute of Space and Aeronautical Science, University of Tokyo, 1980).
15. Banks, P. M. & Kockarts, G. in *Aeronomy* Part B, 170 (Academic, New York, 1973).

C_3H_8 and C_3H_4 in Titan's atmosphere

W. C. Maguire, R. A. Hanel, D. E. Jennings, V. G. Kunde & R. E. Samuelson

Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Four bands of propane C_3H_8 and two of methyl acetylene C_3H_4 have been identified in the Voyager IR spectrum of Titan. Stratospheric abundances of 2×10^{-5} for C_3H_8 and 3×10^{-8} for C_3H_4 have been determined for the mid-latitude region. A feature at 1,154 cm^{-1}, previously assigned solely to CH_3D, is now identified at least in part due to C_3H_8.

THE Voyager 1 IR instrument (IRIS)[1] obtained several hundred spectra of Titan when the angular diameter of the disk was at least five times the angular field of IRIS. Spectra of the centre of the disk, the polar regions, and both limbs all revealed the signatures of many hydrocarbons and of HCN in emission, indicating the presence of these gases in a warm stratosphere. In

0028–0836/81/340683—04$01.00

© 1981 Macmillan Journals Ltd

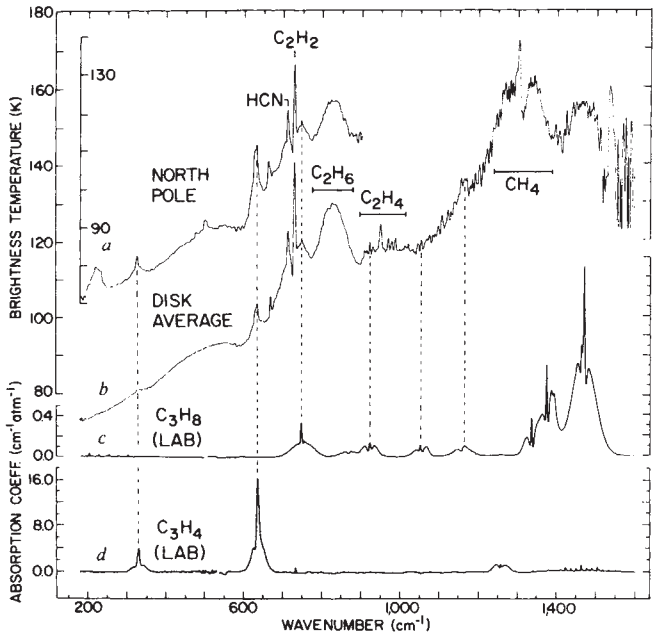


Fig. 1 Comparison of observed Titan spectrum (4.3 cm^{-1} resolution) with laboratory measurements (2 cm^{-1} resolution) at room temperature of C_3H_8 and C_3H_4. *a*, An average of 29 spectra from the north polar region of Titan. *b*, An average of 346 dayside spectra, principally from the centre of the disk but also including data from latitudes as high as 60° N or S. Laboratory spectra of C_3H_8 (*c*) and C_3H_4 (*d*) are also shown.

some cases firm identifications and estimates of the abundances were obtained; in other cases, such as for C_3H_8 and C_3H_4 only tentative identifications were made[2]. We provide here additional spectroscopic evidence for the existence of C_3H_8 and C_3H_4 in Titan's stratosphere and give abundance estimates using our laboratory spectra.

Titan and laboratory spectra

Two Titan spectra, containing common strong features, are shown in Fig. 1: the 'disk average' is the average of 346 individual spectra taken near the centre of the disk at a spatial resolution between 1/5 and ~1/50 of the disk diameter; the average emission angle is ~35°. The signal-to-noise ratio is ~430 at 250 cm^{-1}, ~130 at 800 cm^{-1}, ~16 at 1,000 cm^{-1}, and ~35 at 1,500 cm^{-1}. The 'north pole' spectrum is an average of 29 spectra recorded at 68° latitude with an average emission angle of 63.4°. The signal-to-noise ratio is lower in this spectrum, ~125 at 250 cm^{-1} and 35 at 800 cm^{-1}. The behaviour of the continuum emission between 200 and 600 cm^{-1} is discussed elsewhere[3].

The laboratory spectra were recorded at room temperature in two segments: 200–500 cm^{-1} and 400–1,600 cm^{-1}. For C_3H_8 the pressure of the sample was 994 mbar for the low and 404 mbar for the high frequency region. For C_3H_4, the pressures were 47 mbar for the low and 9.3 mbar for the high frequency region. Sample runs were ratioed with empty cell runs to produce transmission spectra from which absorption coefficients were derived (shown in Fig. 1). Band designations, wavenumbers and band strengths are listed in Table 1 for C_3H_8 and C_3H_4 for the observed 200–1,600 cm^{-1} range. Several combination bands at 1,255 cm^{-1} of C_3H_4 also appear in Fig. 1 in the laboratory spectrum.

Comparison of Titan and laboratory spectra

The average disk spectrum shows weak but definite spectral features of the C_3H_8 fundamentals at 748, 922, 1,054 and 1,158 cm^{-1} (dashed lines in Fig. 1). The *R* branch of C_2H_2 is superimposed on the propane band but the 748 *Q*-branch of C_3H_8 is clearly present. The weaker bands at 922, 1,054 and 1,158 cm^{-1} are all visible although in a spectral region where C_2H_4 and other hydrocarbons have features of comparable strength. We have modelled the bands of CH_4, C_2H_2, C_2H_4, C_2H_6 and CH_3D in the 700–1,400 cm^{-1} range and found that none of these gases can account for the spectral features now assigned to C_3H_8. Comparison of the spectra also shows that a feature at 1,154 cm^{-1}, which has previously been identified[4] as due only to CH_3D, is due principally to C_3H_8. The ν_4 *Q*-branch of CH_3D appears to ride on this feature. The strongest C_3H_8 band in the laboratory spectrum at 1,400 cm^{-1} shown in Fig. 1 shows no corresponding feature in the Titan spectrum, partly because CH_4 and many other hydrocarbons have strong and overlapping bands in this spectral interval, and partly because the emission originates at about the 6 mbar level where thermal contrast is greatly reduced (Fig. 2).

The disk spectrum and to a greater extent the polar spectrum show the fundamentals of C_3H_4 at 328 cm^{-1} and at 633 cm^{-1}. These features are associated with the same gas because they tend to exhibit similar behaviour with position on the disk, decreasing in strength with increasing distance from the north polar region. They also show identical behaviour in scans across the limb (not shown), rising and falling in unison as each scan proceeds from space to well onto the disk. No signatures of the combination bands at 1,200–1,400 cm^{-1} are expected because of their weakness and the above reasons for C_3H_8.

The identification of C_3H_4 is made firmer by comparing the relative strengths of the features at 328 and 633 cm^{-1} in the Titan spectra with those in the laboratory spectra. Figure 2 shows most of the mass of gaseous C_3H_4 must be between the condensation point ($T \sim 110$ K; $P \sim 28$ mbar) and a point one pressure scale height above ($T \sim 150$ K; $P \sim 10$ mbar). Thus the temperature of the bulk of C_3H_4 is ~130 K. However, the peak brightness temperature observed in the spectrum (Fig. 1) is well below this temperature. Because the *Q*-branch width is comparable to the IRIS spectral resolution, it follows that the bands are optically thin, and the *Q*-branch intensity can be approximated by

$$I_\nu = \varepsilon_\nu B_\nu(130\ \mathrm{K}) \qquad (1)$$

Table 1 Fundamental frequencies and IR absorption intensities of C_3H_8 and C_3H_4

Band	Wave number* (cm^{-1})	Intensity† (cm^{-2} atm^{-1} at 300 K)
Propane		
ν_{22}	(223.0)‡	—
ν_9	360.3	0.4 ± 0.4
ν_{21}	748.1	10.1 ± 1.0§
ν_8	860.3	2.2 ± 0.69
ν_{16}	921.7	4.63 ± 0.65
ν_{15}	1,053.8	4.0 ± 0.45
ν_7	1,157.5	5.04 ± 0.61
ν_{20}	1,191.5	
ν_{14}	1,338	10 ± 3
ν_{13}	1,378	24 ± 4
ν_6	1,392	
ν_5	1,462	69.9 ± 7.3
ν_{12}	1,464	
ν_{19}	1,472	
ν_4	1,476	
Methyl acetylene		
ν_{10}	327	67.3 ± 7.0‖
ν_9	633	282 ± 28‖
ν_5	930	4.1 ± 2.0
ν_8	1,052	2 ± 1
ν_4	1,380	6.1 ± 4.1
ν_7	1,452	72.7 ± 6.5

*From ref. 7.
†From ref. 8 for C_3H_8; from ref. 14 for C_3H_4.
‡Estimated from combination bands.
§Our measured value. Kondo and Saëki[8] report 8.9 ± 1.0 for ν_{21}.
‖Our measured values. Kondo and Koga[14] report 67.8 ± 4.1 and 355 ± 18 for ν_{10} and ν_9, respectively.

© 1981 Nature Publishing Group

where the emissivity of the C_3H_4 layer, ε_ν, is the product of the gas density and band strength and B_ν is the Planck function at 130 K. The band intensity ratio is therefore given by

$$\frac{I_{633}}{I_{328}} = \frac{S_{633}B_{633}(130\text{ K})}{S_{328}B_{328}(130\text{ K})} \quad (2)$$

where S is the band intensity. The Planck function ratio in equation (2) is $\sim 1/4$, while, from the laboratory spectrum in Fig. 1, $S_{633}/S_{328} \sim 4$ (see Table 1); thus the calculated value for I_{633}/I_{328} is ~ 1 provided the band strength ratio is approximately independent of temperature between 130 and 273 K. On the other hand the brightness temperature spectrum of 68° N (see Fig. 1, ref. 5) indicates that the observed value of I_{633}/I_{328} *is* ~ 1.2, depending on the exact location of the continuum: this agreement is consistent within the uncertainties involved.

Abundance determinations

For normal viewing, the solution to the equation of transfer over a narrow spectral range $\Delta\nu$ can be written

$$I_{\Delta\nu} \simeq \int_P B_{\Delta\nu} \frac{\partial}{\partial P} \times \left[\int_{\nu'}^{\nu} \phi(\nu - \nu') \exp\left\{ -\int_{P'} \left(\sum_j \sum_i q_j k_{ji}(\nu' - \nu_i) \right) \frac{dP'}{g} \right\} d\nu \right] dP \quad (3)$$

where I is the observed radiance, $\Delta\nu$ is the spectral resolution, ϕ is the instrument convolution function, B is the Planck function, g is the acceleration due to gravity, P is the pressure, and k is the mass absorption coefficient of a gas with a mole fraction q. The summations are over all gases j and all contributing lines i of a single gas. The approximation is very good when $\Delta\nu$ is small because B is essentially constant across such a narrow spectral interval.

Solving for q requires first that the vertical thermal profile of the atmosphere be determined. Second, the vertical distributions and mole fractions of all absorbing gases must be specified; initially the fraction q will be a trial value. Third, the transmission functions of all absorbing gases are calculated. Finally, an iteration is performed on q until equation (3) is satisfied.

We have adopted the shape of the radio occultation profile[6] between 1,600 and 10 mbar, normalizing it to a surface temperature of 94 K. This profile has been extrapolated to 1 mbar using IRIS data covering the 1,250–1,310 cm^{-1} range. Methane, the major emitter in this region, is assumed to be uniformly mixed.

Previous abundance determinations[2] using a line-by-line transmission function program have specified the mixing ratios for C_2H_6, C_2H_2, and HCN. Each gas was assumed to be uniformly mixed down to its condensation level, below which it was reduced to vanishingly small amounts. The cutoff was 50 mbar for C_2H_6 and C_2H_2 and 10 mbar for HCN. To adjust the spectral continuum to the data a stratospheric aerosol was added above 5 mbar, assigned an optical depth of 0.01 at 700 cm^{-1} and made to increase linearly to 0.05 at 1,400 cm^{-1}. The exact location of the aerosol is unimportant as long as it is optically thin[3].

For C_3H_8, a synthetic spectrum was constructed for the ν_{21} band at 748 cm^{-1}. The band intensity of ν_{21} was measured from the 2 cm^{-1} resolution laboratory spectra to be 10.1 ± 1.0 cm^{-2} atm^{-1} at 300 K. Other molecular parameters are from refs 7 and 8. Because the distribution of intensities within a 4.3 cm^{-1} interval is less important to the present analysis than the integrated intensity in this interval, only representative line positions and intensities were used to fit our room-temperature laboratory spectrum. No molecular spectroscopic analysis was performed. Propane was also assumed to be uniformly mixed down to 50 mbar in the model, below which its abundance was set to 0. Equation (3) was iterated until convergence was reached for $q(C_3H_8)$. We infer

$$q(C_3H_8) \sim 2 \times 10^{-5} \quad (4)$$

for the disk average which should be accurate to within a factor of three. The major sources of error involve the imprecise vertical modelling of the other hydrocarbons and the aerosol, and an uncertain 748 cm^{-1} Q-branch strength of C_3H_8. Because the entire 748 cm^{-1} band is weak, small errors in the relatively large background continuum lead to large errors in the inferred C_3H_8 mixing ratio. The concentration of C_3H_8 seems to be fairly constant over the disk.

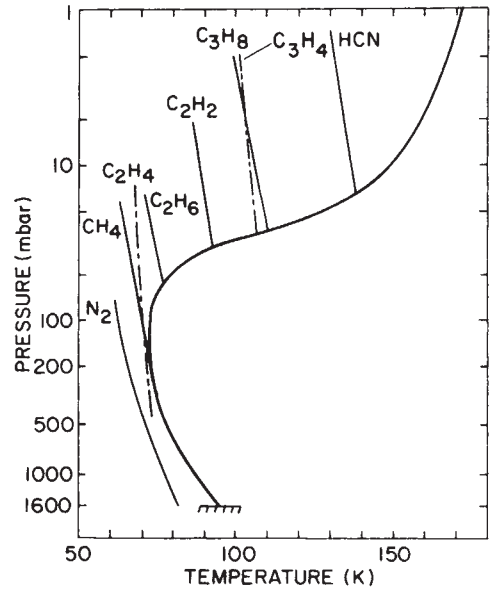


Fig. 2 Vertical temperature profile of Titan's atmosphere[3,6] and the condensation curves of N_2 and several hydrocarbons with mole fractions listed in Table 1 of ref. 5. Uniform mixing is assumed throughout the stratosphere. HCN will condense at about 15 mbar and C_2H_6 at about 60 mbar. C_2H_4 will not condense at all in the assumed concentration. The condensation products are solid particles for HCN, C_3H_4, C_2H_2, C_2H_6 and CH_4 and liquid droplets for C_3H_8. Very little gas will exist below the condensation point because of the steepness of the temperature profile and vapour pressure curves.

We could now determine the mixing ratio of C_3H_4 from equation (3) if we knew the related values of $k(\nu)$ for this gas. Fortunately the integrated ν_9 band intensity and Q-branch intensity at 633 cm^{-1} measured from the laboratory spectra is sufficient for obtaining a good approximation to the mixing ratio when C_3H_4 is optically thin throughout this band, as it is for Titan. We can expand the transmission function in equation (3) in a power series and retain only the first two terms. On inverting the orders of integration and summation in equation (3) we have

$$\int_{\nu'} \phi(\nu - \nu') \exp\left[-\int_P \sum_j \sum_i q_j k_{ji}(\nu' - \nu_i) \frac{dP}{g} \right] d\nu'$$
$$\simeq \int_{\nu'} \phi(\nu - \nu') \left[1 - \frac{1}{g} \int_P \left(\sum_j \sum_i q_j k_{ji}(\nu' - \nu_i) \right) dP \right] d\nu'$$
$$= 1 - \frac{1}{g} \int_P \left[\sum_j \sum_i q_j \left(\int_{\nu'} \phi(\nu - \nu') k_{ji}(\nu' - \nu_i)\, d\nu' \right) \right] dP \quad (5)$$

The second term on the right hand side is the optical depth τ_ν. On integrating over a spectral interval $\delta\nu$ we obtain

$$\delta\nu \tau_{\delta\nu} = -\frac{1}{g} \int_P \left(\sum_j \sum_i q_j k_{ji} \right) dP \quad (6)$$

where k is the integrated strength of an individual line and $\tau_{\delta\nu}$ the optical depth averaged over $\delta\nu$. If we restrict $\delta\nu$ to the width of the Q-branch of a single gas, equation (6) becomes approximately

$$\tau_Q = -\frac{S^Q}{g} \int_P q\, dP \quad (7)$$

where S^Q is the Q-branch strength and τ_Q is the Q-branch optical depth of the gas. Equation (7) is an excellent approximation as long as there is minimal saturation (that is the gas is

© 1981 Nature Publishing Group

everywhere optically thin), and the strength of the Q-branch is insensitive to temperature. The former is true for the 328 and 633 cm^{-1} C_3H_4 bands on Titan, while the latter is true for fundamental bands generally, for which the lower state energies are small. The Q-branch strength of C_2H_2, for example, increases from 356 to only 369 cm^{-2} atm^{-1} over the temperature range 300–150 K. If we restrict solutions of q to those for which there is uniform mixing down to some cutoff level, equation (7) becomes

$$\tau_Q = \frac{P}{g} q \frac{S^Q}{\delta\nu} \tag{8}$$

where P is less than the atmospheric pressure at the condensation level, provisionally assumed to be 50 mbar for C_3H_4.

We could substitute equation (8) back into equation (3) and solve for $q(C_3H_4)$ but because the available line-by-line program cannot readily accept a hybrid combination of transmission functions, the following artifice has been adopted. Available line parameters for the 729 cm^{-1} C_2H_2 Q-branch replace those for the 633 cm^{-1} C_3H_4 Q-branch. Equation (3) is then iterated and a fictitious mole fraction q^* is determined. The corresponding Q-branch optical thickness derived from equation (8) is correct, however, as long as it is small (no saturation), because it was obtained by fitting the calculations to observed data. Equations (5)–(8) demonstrate that the Q-branch optical thickness is independent of the number, strength, and way in which the individual lines are distributed in the band; q is adjusted accordingly. Hence for optically thin layers qS^Q is an invariant which depends only on the observations and the background opacity. This enables us to obtain the actual C_3H_4 abundance from the expression

$$q(C_3H_4) = \frac{S^Q(C_2H_2)}{S^Q(C_3H_4)} q^* \tag{9}$$

where $S^Q(C_3H_4)$ is obtained from our laboratory data, and $q^*S^Q(C_2H_2)$ is inferred from the strength of the 633 cm^{-1} C_3H_4 Q-branch in the Titan spectrum. We find

$$q(C_3H_4) \sim 3 \times 10^{-8} \tag{10}$$

for the disk average with an uncertainty of a factor of ~3 due to attendant uncertainties in modelling the vertical distribution and placement of the continuum, as well as the treatment of the overlapping C_4H_2 Q-branch at 628 cm^{-1} (ref. 5). The reasons for the enhancement in the north are not known; one possible explanation is that the differences are associated with the north polar hood. The atmospheric hydrocarbon composition of Titan determined from Voyager IRIS data is summarized in Table 1 of ref. 5 for the disk average.

Organic chemistry is significantly different on Titan compared with that on Jupiter or Saturn partly because of a different photo- and high-energy particle radiation chemistry due to a relatively lower abundance of H_2 and higher abundances of N_2 and CH_4 on Titan[3], and partly because of the presence of an atmospheric sink at Titan's surface[9–14].

We thank J. Heaney and K. Stewart for operating the laboratory Fourier transform spectrometer, and J. Faris for laboratory assistance. This work is supported in part by the NASA Planetary Atmospheres program. The laboratory interferometer is operated by the Optics Branch, Goddard Space Flight Center.

Received 3 June; accepted 20 July 1981.

1. Hanel, R. *et al. Appl. Opt.* **19**, 1391–1400 (1980).
2. Hanel, R. *et al. Science* **212**, 192–200 (1981).
3. Samuelson, R. E., Hanel, R. A., Kunde, V. G. & Maguire, W. C. *Nature* **292**, 688–693 (1981).
4. Gillett, F. C. *Astrophys. J. Lett.* **201**, L41–L43 (1975).
5. Kunde, V. G. *et al. Nature* **292**, 686–688 (1981).
6. Tyler, G. L. *et al. Science* **212**, 201–206 (1981).
7. Gayles, J. N. & King, W. T. *Spectr. Acta* **21**, 543–557 (1965).
8. Kondo, S. & Saëki, S. *Spectr. Acta* **29A**, 735–751 (1973).
9. Strobel, D. F. *Icarus* **21**, 466–470 (1974).
10. Strobel, D. F. *J. atmos. Sci.* **30**, 489–498 (1973).
11. Strobel, D. F. *Rev. Geophys. Space Phys.* **13**, 372–382 (1975).
12. Capone, L. A., Dubach, J., Whitten, R. C., Prasad, S. S. & Santhanam, K. *Icarus* **44**, 72–84 (1980).
13. Chang, S., Scattergood, T., Aronowitz, S. & Flores, J. *Rev. Geophys. Space Phys.* **17**, 1923–1933 (1979).
14. Kondo, S. & Koga, Y. *J. chem. Phys.* **69**, 4022–4031 (1978).

C_4H_2, HC_3N and C_2N_2 in Titan's atmosphere

V. G. Kunde, A. C. Aikin, R. A. Hanel, D. E. Jennings, W. C. Maguire & R. E. Samuelson

Goddard Space Flight Center, Code 693.2, Greenbelt, Maryland 20771, USA

The compounds C_4H_2, HC_3N, and C_2N_2 have been detected in trace amounts in the stratosphere of Titan. The identification of two compounds containing nitrogen, in addition to HCN, provides further evidence for the abundance of free N_2 on Titan.

THE Titan atmosphere is characterized by hydrocarbon and nitrogen-containing organic compounds in an atmosphere which near the surface is predominantly composed of molecular nitrogen, methane, molecular hydrogen and possibly argon[1]. Molecules containing oxygen have not been detected. Many organic compounds have now been identified in the stratosphere of Titan from ground-based and spacecraft IR spectroscopic observations. The first hydrocarbon identification was the ground-based observation of methane[2] (CH_4). Subsequent ground-based measurements identified ethane[3] (C_2H_6), acetylene[4] (C_2H_2) and ethylene[5] (C_2H_4). Strong IR emissions due to these gases have also been observed with the IR instrument (IRIS)[6] on Voyager 1. The IRIS spectrum also indicates the presence of hydrogen cyanide (HCN), methyl acetylene (C_3H_4), and propane[6,7] (C_3H_8). We now identify three additional organic compounds: diacetylene (C_4H_2), cyanoacetylene (HC_3N), and cyanogen (C_2N_2).

Observations

Voyager 1 took IR measurements of the atmosphere of Titan from ~8 h before to ~3 h after closest approach. An average of 346 spectra, mostly from the centre of the disk but including data from latitudes as high as ±60°, is shown in Fig. 1. The average emission angle is ~35°. Strong emission features associated with stratospheric gases appear above a background continuum (see Fig. 1 and Table 1 which also indicate weaker emission features due to C_3H_4 and C_3H_8). The high latitude spectrum is an average of 30 individual spectra obtained at 68° N latitude near or on the dark polar hood observed in Voyager images[8]. The average emission angle is ~63°. Several features are evident in the 200–700 cm^{-1} region of the 68° N spectrum but not in the midlatitude disk average (220, 233, 500, 663 cm^{-1}). Other features are enhanced in the high polar region (328, 628, 633 cm^{-1}). Some enhancement is caused by the higher emission angle of the north polar spectrum compared

0028–0836/81/340686—03$01.00

© 1981 Macmillan Journals Ltd

with that of the spectra recorded further south. However, in view of the low stratospheric temperatures in the north, the strong enhancement of some spectral features implies a genuine enrichment of these species at high northern latitudes. Differences in temperatures and chemical processes associated with the polar hood may be responsible for the enrichment.

The limb spectrum shown in Fig. 2 is an average of three individual spectra taken at grazing incidence above the north polar cap. The tangent ray defining the centre of the instrument field of view passes ~330 km above the surface of Titan at its closest point, while the diameter of the field of view at this point is ~190 km. Thus, there is considerable space contamination in the spectrum, which depresses the background continuum relative to the sharp emission features producing clear separation between the 220 and 233 cm^{-1} features.

Identification of C_4H_2, HC_3N and C_2N_2

After accounting for the emission features of C_3H_4, HCN, C_2H_2, C_3H_8, C_2H_6, C_2H_4, and CH_4, several prominent features still remained unidentified in the 200–700 cm^{-1} region of the IRIS spectra. The sharp appearance of these features (220, 233, 500, 628, 633 cm^{-1}) suggested association with *Q* branch absorptions of minor stratospheric constituents.

The assignment of these five unidentified spectral features to C_4H_2, HC_3N and C_2N_2 was based on: (1) a correspondence in wavenumber and in the shape of the emission feature with laboratory spectra of the candidate molecules; (2) the apparent agreement of the relative intensities of the features with the corresponding strengths derived from the laboratory spectra if the candidate molecule has more than one feature in the observed spectrum. We only investigated gases which were considered possible constituents of the atmosphere of Titan. A list of potential constituents was assembled from organic chemistry studies involving photochemical and energetic particle bombardment processes[9–11]. Spectra in the 200–1,400 cm^{-1} region were obtained for these potential atmospheric constituents at nearly the same spectral resolution as Voyager IRIS. (The laboratory spectra were obtained at 2 and 4 cm^{-1} resolution with a Fourier transform spectrometer operated by the Optics Branch, Goddard Space Flight Center. Samples were contained in a 3-cm length cell at pressures ranging from 11 to 140 torr. Sample temperatures were ambient (300 K).)

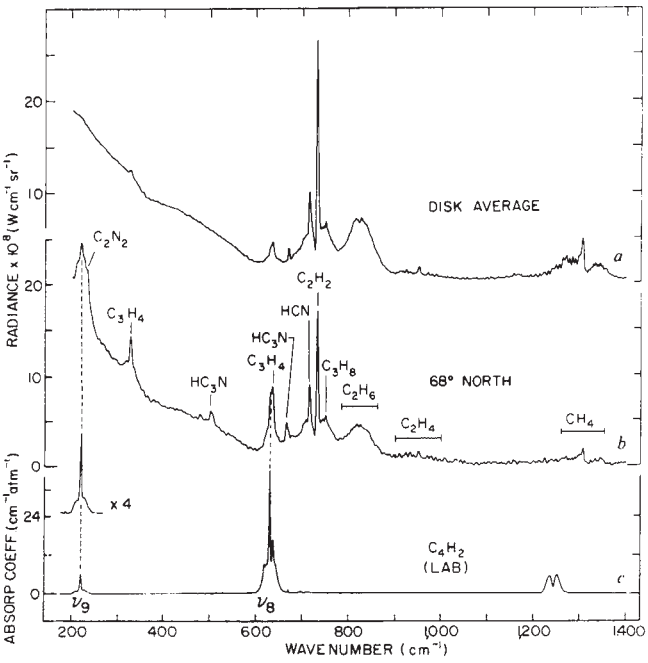


Fig. 1 Comparison of laboratory and Titan radiance spectra. The disk average spectrum (*a*) represents mainly midlatitude regions on Titan while the 68° N spectrum (*b*) is of a region near or over the northern polar cap. Several emission features in the 200–700 cm^{-1} region are enhanced in the 68° N spectrum. *c*, A laboratory spectrum for diacetylene (C_4H_2)

Table 1 Organic compounds in Titan's stratosphere inferred from Voyager IRIS

Chemical family	Gas	Wave number (cm^{-1})	Approximate mole fraction
Carbon–hydrogen			
	CH_4	1304	3×10^{-2}
	C_2H_6	822	2×10^{-5}
	C_3H_8	748	2×10^{-8}
	C_2H_2	729	2×10^{-6}
	C_2H_4	950	4×10^{-7}
	C_3H_4	325,633	3×10^{-8}
	C_4H_2	220,628	10^{-8}–10^{-7}
Carbon–hydrogen–nitrogen			
	HCN	712	2×10^{-7}
	HC_3N	500, 663	10^{-8}–10^{-7}
Carbon–nitrogen			
	C_2N_2	233	10^{-8}–10^{-7}

C_4H_2 is expected to be present in the upper atmosphere of Titan as a result of methane photolysis and hydrogen escape, leading to C_2H and to the production of C_4H_2 by the reaction $C_2H + C_2H_2$. The two features in the IRIS spectrum at 220 and 628 cm^{-1} correspond to the *Q* branches of the ν_9 and ν_8 fundamental vibration–rotation bands of C_4H_2, respectively[12]. These two features are marked in Fig. 1 but may be seen most distinctly in the high air mass spectrum of Fig. 2. The 220 cm^{-1} C_4H_2 feature is partially overlapped by the 233 cm^{-1} C_2N_2 emission and the 628 cm^{-1} feature by the C_3H_4 emission at 633 cm^{-1}. However, the 4.3 cm^{-1} resolution allows identification of C_4H_2. The intensity ratio for the two C_4H_2 features in the observed spectrum may be approximated by[7]

$$\frac{I_{220\ \mathrm{cm}^{-1}}}{I_{628\ \mathrm{cm}^{-1}}} = \frac{S^{Q}_{220}}{S^{Q}_{628}}\,\frac{B_{220}(130\ \mathrm{K})}{B_{628}(130\ \mathrm{K})}$$

where *S* is the strength of the *Q* branch of the molecular band, and *B* is the Planck function at 130 K. *Q* branch strengths derived from laboratory data yield an intensity ratio of ~0.74, which agrees with the observed intensity ratio of ~0.72 for the north limb spectrum (Fig. 2). The observed intensity ratio is therefore consistent with the C_4H_2 identification. The C_4H_2 band at 1,250 cm^{-1} is too weak to appear in the Titan spectrum.

HC_3N is expected in Titan's atmosphere from the reaction of CN with C_2H_2 and C_2H_4, where the CN is produced from the dissociation of HCN and the C_2H_2 from methane photolysis[9]. As indicated by comparison with the laboratory spectrum in Fig. 2, the HC_3N emissions from the ν_6 and ν_5 fundamental vibration–rotation bands[13,14] at 500 and 663 cm^{-1}, respectively, are easily identified. The intensity ratio of the two features ($I_{500\ \mathrm{cm}^{-1}}/I_{663\ \mathrm{cm}^{-1}}$) are approximately the same in the observed and the laboratory spectra (~0.8), thus strengthening the HC_3N identification. The HC_3N band at 1,310 cm^{-1} is too weak to appear in the Titan spectrum.

C_2N_2 is predicted to occur in the Titan atmosphere through the combination of a CN radical, either with HCN or with another CN radical[9]. The only C_2N_2 emission feature occurring in the IRIS spectrum (Fig. 2) is the strong ν_5 fundamental band[15] at 233 cm^{-1}. The detection of C_2N_2 and HC_3N is consistent with experiments carried out for CH_4–N_2 mixtures[16].

The strong emission features of the three gases in the limb spectrum (Fig. 2) indicate that a large fraction of their mass is in the upper stratosphere ($p < 20$ mbar, $T > 130$ K). Additionally large concentrations are not expected to occur in the low temperature environment of the lower stratosphere due to condensation (see Fig. 2 of ref. 7). Thus we can associate the formation of these gases with photochemistry in the upper stratosphere. Crude abundance estimates for these three gases yield mole fractions between 10^{-8} and 10^{-7}, consistent with production by nitrogen ion and methane photolysis photochemistry[9].

Discussion and conclusion

The organic compounds observed in the atmosphere of Titan are summarized in Table 1, which also indicates the approximate

© 1981 Nature Publishing Group

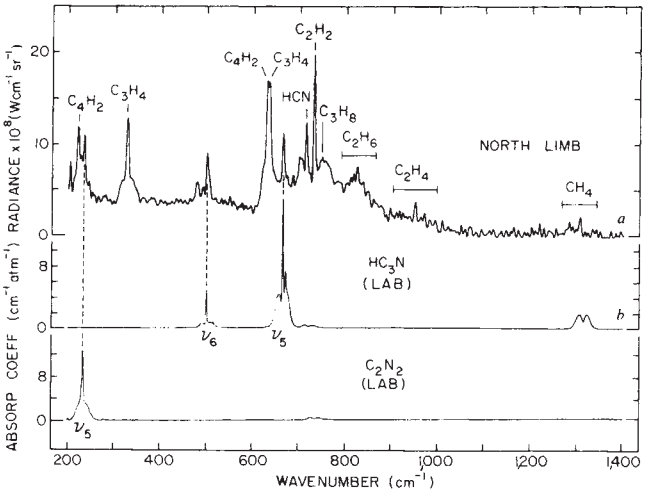


Fig. 2 Comparison of laboratory and Titan radiance spectra. The observed spectrum (*a*) was obtained off the limb of the planet in the northern polar cap region. The high air mass along the line of sight enhances the weak emission features, while space contamination in the field of view depresses the adjacent continuum. Laboratory spectra for cyanoacetylene (HC_3N) (*b*) and cyanogen (C_2N_2) (*c*) are shown.

mole fraction for each compound identified previously, as determined from analysis of the disk average spectrum. These compounds originate by reactions of methane and nitrogen radicals in a predominantly nitrogen atmosphere. Although Titan's location at 9.35 AU reduces the solar flux by a factor of 87 compared with the flux at the Earth, dissociation of significant quantities of methane still occurs due to Lyα radiation. Titan is located within the Saturn magnetosphere and lacks a shielding magnetic field. This causes energetic electrons and protons to ionize and dissociate the principal atmospheric constituents. Bremsstrahlung extends the effective penetration depth of energetic electrons down to an altitude of 130 km. In addition to charged particles from Saturn's magnetosphere Titan is bombarded by galactic cosmic radiation, consisting mostly of protons in the GeV range, which penetrate to the surface. Particle ionization is expected to lead to neutral fragments and ions, including CH, N, CH_2, N_2^+, N^+, CH_4^+, and CH_3. Subsequent ion–molecule reactions involving methane, molecular hydrogen, and molecular nitrogen lead to complex neutral hydrocarbons. The various energy sources, acting over a long period of time, have formed a complex array of hydrogen–carbon–nitrogen compounds.

The wide variety of organic compounds, with fairly high abundances of C_2H_2 and C_3H_8, indicates that the chemistry of Titan is significantly different from that of Jupiter and Saturn. One reason is the low H_2 abundance ($\sim$0.2%)[1] which prevents rapid recycling of CH_4 (ref. 18), and yields higher relative levels of the methane photolysis reactive radicals (CH_2, CH_3, C_2H). Also, these and the CN radical from the dissociation of HCN, do not react with the major constituent N_2 and thus are available to react with minor hydrocarbon species to form more complex compounds. Additionally, the reduced solar flux leads to an increased lifetime of minor atmospheric constituents against photodissociation.

Many of the Titan constituents have been detected in interstellar sources[18–20]. Among those with non-zero dipole moments (HCN, HC_3N, C_3H_4, and C_3H_8) only C_3H_8 has not been detected at radio wavelengths for the interstellar sources.

We thank J. Heaney and K. Stewart for operating the laboratory spectrometer, and J. Faris for laboratory assistance. The diacetylene sample was supplied by H. Okabe. Partial support was provided by the NASA Planetary Atmospheres Program.

Received 3 June; accepted 20 July 1981.

1. Samuelson, R. E., Hanel, R. A., Kunde, V. G. & Maguire, W. C. *Nature* **292**, 688–693 (1981).
2. Kuiper, G. P. *Astrophys. J.* **100**, 378–383 (1944).
3. Gillett, F. C. & Forrest, W. J. *Astrophys. J. Lett.* **184**, L93–L95 (1973).
4. Tokunaga, A. *Bull. Am. astr. Soc.* **12**, 669 (1980).
5. Gillett, F. C. *Astrophys. J. Lett.* **201**, L41–L43 (1975).
6. Hanel, R. *et al. Science* **212**, 192–200 (1981).
7. Maguire, W. C., Hanel, R. A., Jennings, D. E., Kunde, V. G. & Samuelson, R. E. *Nature* **292**, 683–686 (1981).
8. Smith, B. A. *et al. Science* **212**, 163–190 (1981).
9. Strobel, D. F. *Geophys. Res. Lett.* (submitted).
10. Chang, S., Scattergood, T., Aronowitz, S. & Flores, J. *Rev. Geophys. Space Phy.* **17**, 1923–1933 (1979).
11. Allen, M., Pinto, J. P. & Yung, Y. L. *Astrophys. J. Lett.* (in the press).
12. Hardwick, J. L. & Ramsay, D. A. *J. molec. Spectrosc.* **76**, 492–505 (1979).
13. Job, V. A. & King, G. W. *Can. J. Chem.* **41**, 3132–3133 (1963).
14. Mallinson, P. D. & Fayt, A. *Molec. Phys.* **32**, 473–485 (1976).
15. Jones, L. H. *J. Molec. Spectrosc.* **45**, 55–64 (1973).
16. Toupance, G., Raulin, F. & Buvet, R. *Orig. Life* **6**, 83–90 (1975).
17. Strobel, D. F. *Icarus* **21**, 466–470 (1974).
18. Lovas, F. J., Snyder, L. E. & Johnson, D. R. *Astrophys. J. Suppl.* **41**, 451–480 (1979).
19. Hall, D. N. B. & Ridgway, S. T. *Nature* **273**, 281–282 (1978).
20. Betz, A. L. *Astrophys. J. Lett.* **244**, L103–L105 (1981).

Mean molecular weight and hydrogen abundance of Titan's atmosphere

R. E. Samuelson, R. A. Hanel, V. G. Kunde & W. C. Maguire

Goddard Space Flight Center, Code 693.2, Greenbelt, Maryland 20771, USA

The 200–600 cm^{-1} continuum opacity in the troposphere and lower stratosphere of Titan is inferred from thermal emission spectra from the Voyager 1 IR spectrometer (IRIS). The surface temperature and mean molecular weight are 94 K $< T_G <$ 97 K and 28.3 $< M <$ 29.2 AMU, respectively. The mole fraction of molecular hydrogen is 0.002 $\pm$ 0.001, which is equivalent to an abundance of $\sim$0.2 $\pm$ 0.1 km amagat.

THERMAL inversion in Titan's atmosphere has been demonstrated by narrow band IR photometry[1,2] and spectrophotometric data[3] which showed that at least methane, ethane and probably ethene exist in a warm stratosphere overlying a colder background. Danielson *et al.*[4] and Caldwell[5] quantitatively explained the major features of the thermal emission spectrum using simple inversion models.

Models with a deeper atmosphere had also been suggested. But even in the model proposed by Hunten[6], which was remarkably accurate in its thermal and compositional predictions, it was accepted that methane clouds and photochemically induced smog would obscure the atmosphere below the tropopause. McCarthy *et al.*[7] interpreted their data between 330 and 625 cm^{-1} as being consistent with an inversion model. An optically opaque base, either the surface or dense clouds, was assumed to be roughly at a temperature of 74 K, below which little, if any, information was available.

This situation changed when Voyager 1 IR data showed limb darkening over the 200–600 cm^{-1} spectral range[8]. The radiance at small emission angles (the angle at the surface between the direction of viewing and the surface normal) was found to be larger than that for large angles near the limb. This could occur

0028–0836/81/340688—06$01.00

© 1981 Macmillan Journals Ltd

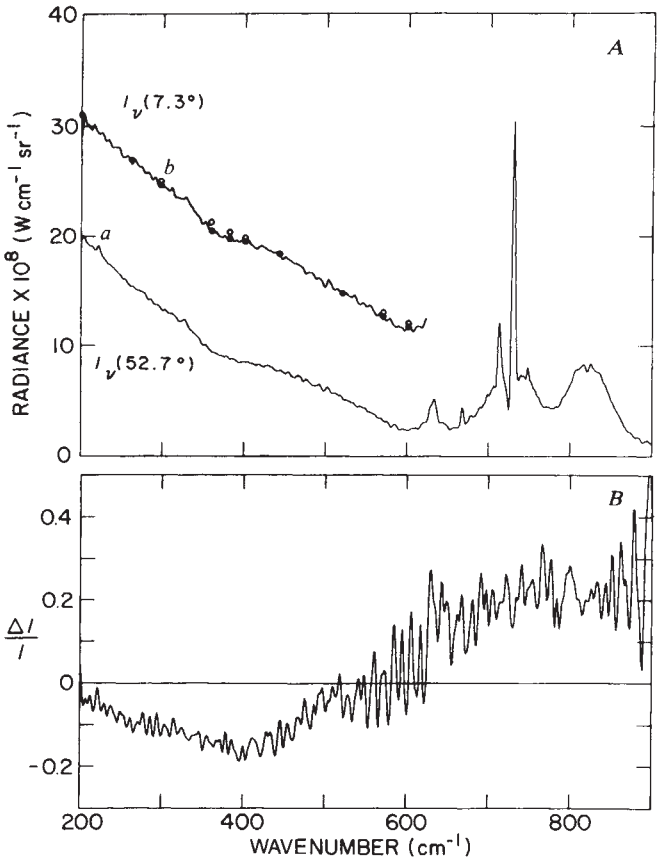


Fig. 1 *A*. *a*, average of 29 spectra near the daytime east limb. The average emission angle is 52.7° and the mean latitude is 7.2 °N. *b*, Average of 10 spectra near the sub-spacecraft point. The average emission angle is 7.3° and the mean latitude is 8.3 °N. Theoretical fits for a hydrogen mole fraction $q = 0.002$ (●); $q = 0$ (○). Values are displaced 10 radiance units upwards for clarity. *B*, northern low-latitude spectral limb function: $\Delta I/I \approx 1 - I_\nu(7.3°)/I_\nu(52.7°)$.

only if the observed radiation originated principally below the temperature minimum, from the troposphere.

The observed spectral limb function is

$$\frac{\Delta I}{I} = \frac{I_\nu(52.7°) - I_\nu(7.3°)}{I_\nu(52.7°)} \quad (1)$$

where $I_\nu(\theta)$ is the radiance at wavenumber ν and emission angle θ. $I_\nu(7.3°)$ and $I_\nu(52.7°)$ are averages of 10 and 29 spectra, respectively, and are shown in Fig. 1 with the spectral limb function.

We have found that at low latitudes the limb function is symmetric about the sub-spacecraft meridian, indicating a lack of diurnal and longitudinal variation, although Flasar *et al.*[9] find evidence for a slight latitudinal inhomogeneity at ~10° south latitude where a weak opacity transition at 200 cm^{-1} seems to occur. This coincides approximately with the visual brightness transition found near the equator[10]. Thus the limb function given by equation (1) should be free of horizontal inhomogeneity as the selected data sets are associated with fields of view that are small, closely grouped, and are between 0° and 15° N latitude. Because this latitude range also includes the ingress Earth occultation point, we can incorporate the radio science temperature profile[11] directly into a quantitative interpretation of our data.

Models

The two spectral averages indicated in equation (1) yield independent information, principally because the large difference in emission angles ensures that different effective emission levels in the atmosphere are being sampled. Thus at each wavenumber two independent parameters characterizing the radiative properties of the atmosphere can be inferred. We have chosen for these the optical thicknesses of the stratosphere $\Delta\tau_S$ and tropopause $\Delta\tau_T$.

We restrict the computational models to the range 200–600 cm^{-1}. Solutions for higher wavenumbers do not converge, probably because of insufficient signal below the tropopause. A simple exponential law of absorption is assumed, which is adequate as long as clouds, hazes and pressure-induced absorption by gases are the only opacity sources. Thus we avoid the spectral ranges 205–255 cm^{-1}, 305–355 cm^{-1}, 460–515 cm^{-1} and above 605 cm^{-1}, where sharp-line spectral features are seen in the spectra[8]. The formal solution to the equation of transfer

$$I_\nu = \int_0^\infty B_\nu(\tau)\, e^{-\tau/\mu} \frac{d\tau}{\mu} \quad (2)$$

can be expressed numerically by

$$I_\nu(\mu_k) = \sum_{i=1}^{n} B_\nu(\tau_i)\left\{\exp\left[-\sum_{j=i+1}^{n} \frac{C_j}{\mu_k}\Delta t_j\right]\right\} \times \left\{1 - \exp\left[-\frac{C_i}{\mu_k}\Delta t_i\right]\right\} (k = 1, 2) \quad (3)$$

where $B_\nu(\tau_i)$ is the Planck function at wavenumber ν (constant across layer i), τ is the optical depth, μ_k the direction cosine of the kth emission angle, n the number of layers (including the ground), and

$$\Delta\tau_i = C_i \Delta t_i (i = 1, \ldots, n) \quad (4)$$

is the optical thickness of layer i. The relative vertical distribution of opacity is specified in the model by the reduced optical thickness Δt_i. The C_i in equations (3) and (4) are then solved by an iterative technique. For each wavenumber interval the C_i provides multiplication factors which convert the assumed relative optical thicknesses, Δt_i, into absolute values $\Delta\tau_i$. Because we have only two free parameters available we assign one value to the stratosphere and the other to a layer at the tropopause; the latter is about the level where we expect the upper boundary of methane clouds. The height of the stratospheric layer as well as its vertical extent is varied from case to case. In some cases a third multiplier $C = 1$ is specified in the troposphere to accommodate pressure-induced absorption.

The temperature profiles are patterned after the radio occultation ingress profile (ref. 11 and G. L. Tyler and G. Lindal,

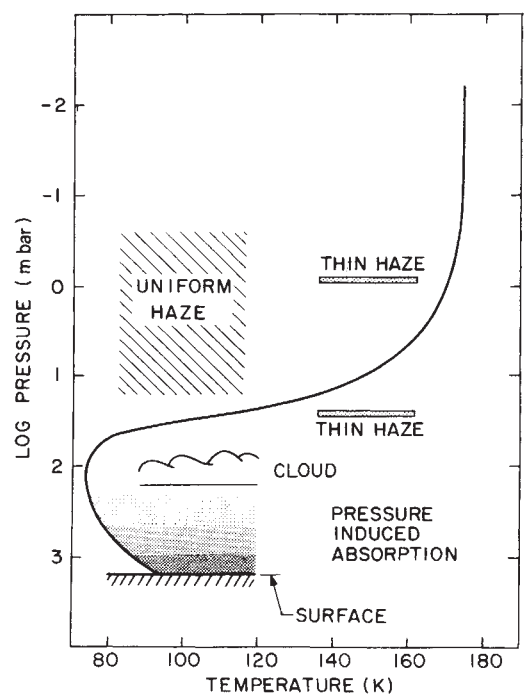


Fig. 2 Temperature profile corresponding to our nominal model. The shape between 10 and 1,600 mbar is taken from data provided by G. L. Tyler and G. Lindal and normalized to a surface temperature $T_G = 95$ K. The extrapolation above 10 mbar is forced to be consistent with brightness temperatures observed in the 1,304 cm^{-1} ν_4 band of methane in the IRIS spectra.

© 1981 Nature Publishing Group

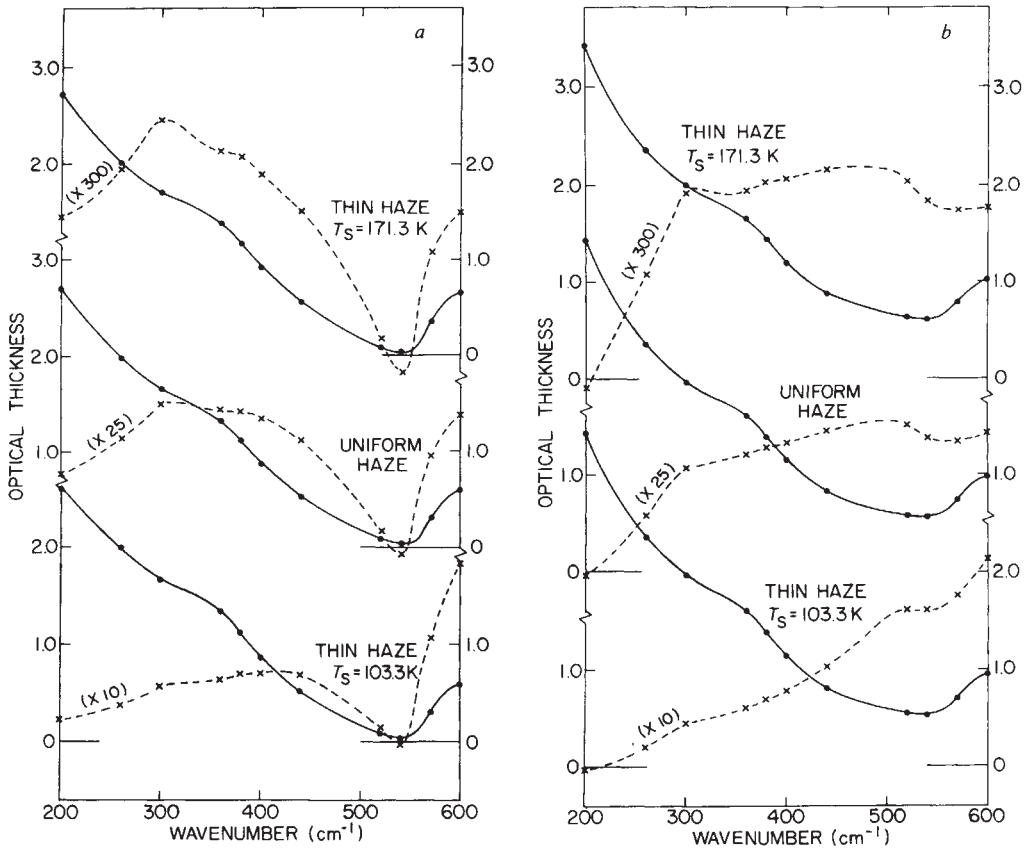


Fig. 3 Optical thicknesses of the stratosphere $\Delta\tau_S$ (– – × – – ×) and tropopause $\Delta\tau_T$ (—●—). The three models for $\Delta\tau_S$ are expanded by factors of 300, 25, and 10 for clarification. Surface temperature: *a*, $T_G = 94$ K; *b*, $T_G = 96.3$ K.

personal communication). The shape of this profile is retained, but the magnitude is normalized to different values of the surface temperature T_G, which is equivalent to assuming different (constant with altitude) mean molecular weights M. Above 10 mbar the profile is extended to 0.01 mbar by extrapolating to 174 K, an asymptotically limiting value consistent with the brightness temperatures at the Q-branch of CH_4 at 1,304 cm^{-1} in the low-latitude IRIS spectra. Partial checks using the CH_4 P-branch brightness temperatures at 1,260 cm^{-1} show internal consistency to within ~3 K at the 5 mbar level. A nominal working profile with $T_G = 95$ K is shown in Fig. 2.

Figure 2 also shows the various forms of opacity considered in our models. The first set of models consists of one cloud layer at the tropopause and three opacity alternatives in the stratosphere: two thin haze layers located at temperatures of 171.3 K and 103.3 K, and a haze mixed uniformly throughout the stratosphere with an opacity directly proportional to the atmospheric density. Each alternative is computed separately using the radiances measured at both emission angles. The solutions are the total stratospheric and tropopause optical thicknesses $\Delta\tau_S$ and $\Delta\tau_T$ as functions of wavenumber, and are shown in Fig. 3 for surface temperatures of 94 K and 96.3 K. These temperatures are close to the minimum and maximum permissible values, indicating that the range of possible opacity models is bounded by those shown in Fig. 3.

Thermal opacity

Several conclusions can be inferred from Fig. 3. First, although the wavenumber dependence is not well defined, the stratospheric opacity tends to be small at all wavenumbers for all models. Thus the contribution of emission from the stratosphere is approximately

$$I_S = B_\nu \Delta\tau_S \qquad (5)$$

where B_ν is the Planck function associated with layer $\Delta\tau_S$. (For the uniform haze models an appropriate weighting of the Planck function must be carried out.) Because the Planck function corresponding to stratospheric temperatures is large relative to values corresponding to temperatures in the troposphere (see Fig. 2), the stratospheric contribution to the measured radiance is not negligible despite small values of $\Delta\tau_S$, but can amount to a large fraction of the total signal. However, without an independent determination of the vertical distribution of the stratospheric haze, it is not possible to separate B_ν and $\Delta\tau_S$. Thus, because the wavenumber dependence of B_ν varies strongly over the range of stratospheric temperatures, a correspondingly large uncertainty also exists in the wavenumber dependence of $\Delta\tau_S$.

Conversely, the qualitative wavenumber dependence of the tropopause opacity is well defined independently of the way in which the stratosphere is modelled, principally because the stratospheric opacity is too small to impede or modify radiation passing through it from below. From 200 cm^{-1} to ~540 cm^{-1} the tropopause opacity decreases monotonically with a reversing trend observed between 540 and 600 cm^{-1}. Beyond 600 cm^{-1} stable solutions are impossible, apparently because there molecular emission from the stratosphere dominates the spectrum.

The general increase of tropopause opacity with decreasing wavenumber below about 540 cm^{-1} has led to two suggestions for the absorption mechanism[8]: pressure-induced absorption by various CH_4 and N_2 combinations, and absorption by methane clouds. Calculations by R. Courtin (personal communication) imply that pressure-induced absorption could supply only a few per cent of the required opacity at the low temperatures prevailing in the Titan troposphere. Other data[12,13] suggest CH_4 clouds to be a realistic opacity source. The observed thermal dis-

Table 1 Possible mole fractions of different gases at Titan's surface

Gas	Mole fraction
N_2	0.822
Ar	0.116
CH_4	0.060
H_2	0.002

The nominal model has a surface temperature of 95 K and a mean molecular weight of 28.6 AMU. The methane fraction of saturation is slightly less than 0.5. Argon (M = 40 AMU) is postulated to be present, partly because its presence is expected on the basis of its cosmic abundance and partly because it is one of the few gases that will not freeze out at the abundance required to balance the mean molecular weight.

©1981Nature Publishing Group

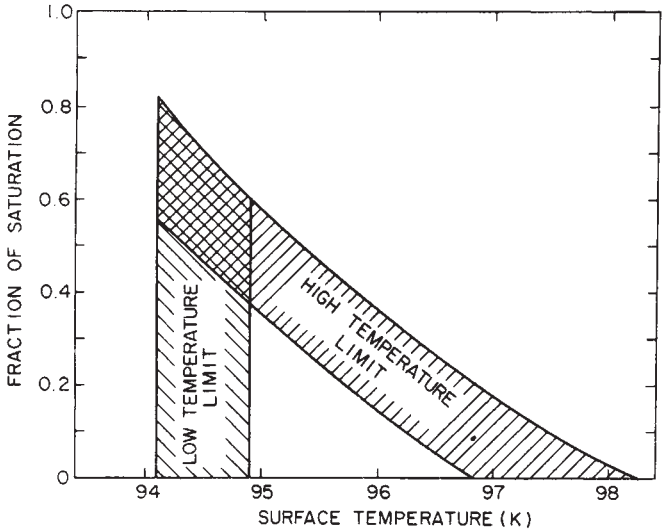


Fig. 4 Minimum and maximum allowable surface temperatures for a given fraction of methane saturation. For a constant methane mole fraction, a fractional saturation of 0.19 at the surface corresponds to complete saturation just below the tropopause. The widths of the bands for the high and low temperature limits give a measure of the uncertainties.

continuity across the equatorial region at 200 cm^{-1} also implies that clouds contribute at least part of the opacity[9]. Until more extensive laboratory data on liquid and solid CH_4 are considered we can only suggest that methane clouds may provide the dominant opacity in Titan's upper troposphere.

Surface temperature and mean molecular weight

At 540 cm^{-1}, the models in Fig. 3*a* show a negative stratospheric opacity and a tropopause opacity that is almost zero; hence the total opacity in the atmosphere is effectively zero, and to lower the surface temperature T_G below 94 K would require both opacities to become negative. Because this is physically impossible, 94 K is a lower limit, and the observed emission at 540 cm^{-1} is dominantly from the surface. If the actual surface temperature were lower, stratospheric emission would have to compensate to yield the measured radiance. This would cause limb brightening, contrary to what is shown in Fig. 1. On the other hand, increasing T_G to 96.3 K causes the calculated stratospheric opacity at 200 cm^{-1} to become negative (Fig. 3*b*). This is a consequence of requiring the minimum atmospheric temperature to be higher than the observed brightness temperature of 74 K at 200 cm^{-1}, again a physical impossibility.

Thus the surface temperature seems to be bounded between close limits, although a few corrections need to be made. Some opacity due to hydrogen absorption is required at 540 cm^{-1} and our modelling suggests that this can be achieved by raising the minimum surface temperature limit to 94.5 ± 0.4 K. Noise in the data and uncertainties in the H_2 absorption coefficients account for the uncertainty in T_G.

The upper limit to the surface temperature is more uncertain. Its assessment depends first on determining an upper limit to the tropopause temperature, and then scaling this to the surface with the aid of the radio occultation profile[11]. If the limb darkening of about 0.8 K observed at 200 cm^{-1} is due partly to multiple scattering in methane clouds, the tropopause temperature must be raised some fraction of this amount. Alternatively, if the 200 cm^{-1} brightness temperature discontinuity of ~1 K across the equator[9] is due to an opacity gradient rather than a horizontal thermal gradient, the tropopause temperature must be reduced accordingly. The upper limit to the temperature at the tropopause becomes 74 K with an estimated uncertainty of ±0.5 K. This scales to a surface temperature upper limit of 96.3 ± 0.7 K.

A further complication is introduced when a variable CH_4 mole fraction below the tropopause is considered, requiring that the mean molecular weight M also be scaled. From published data[14] it is found that the condensation profile for methane becomes tangent to the temperature profile at $T = 75$ K for a methane mole fraction of $q(CH_4) = 0.0266$. Thus, if methane saturates anywhere it will do so at the 75 K level, resulting in a constant mole fraction of 0.0266 above this level and a larger variable fraction below it. Using the T/M value of 2.55 K $(AMU)^{-1}$ at the tropopause[11], we find $M = 29.0$ at $T_T = 74$ K. This leads to a height-independent molecular weight of 29.4 AMU for the remainder of the atmosphere after the contribution from CH_4 is subtracted out. On adopting the radio

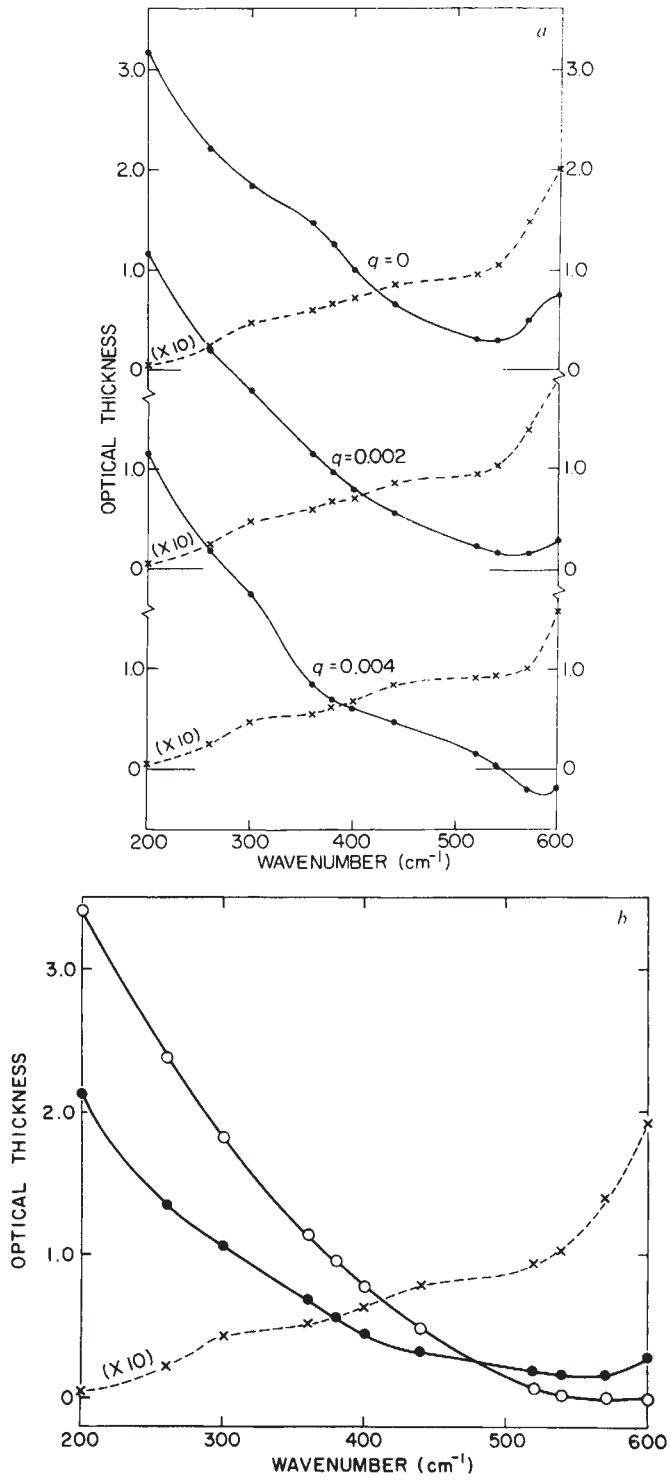


Fig. 5 *a*, Optical thicknesses of the stratosphere $\Delta\tau_S$ (--×--×) and tropopause $\Delta\tau_T$ (—●—) for different H_2 mole fractions q. $\Delta\tau_S$ is restricted to a thin layer with a temperature of 103.3 K, expanded by a factor of 10 for clarification. The surface temperature is $T_G = 95$ K. *b*, $\Delta\tau_S$ and $\Delta\tau_T$ as in *a* with $q = 0.002$. The opacity of an additional absorber (—○—) is proportional to the square of the pressure, thereby restricting its effect mainly to the lower troposphere.

© 1981 Nature Publishing Group

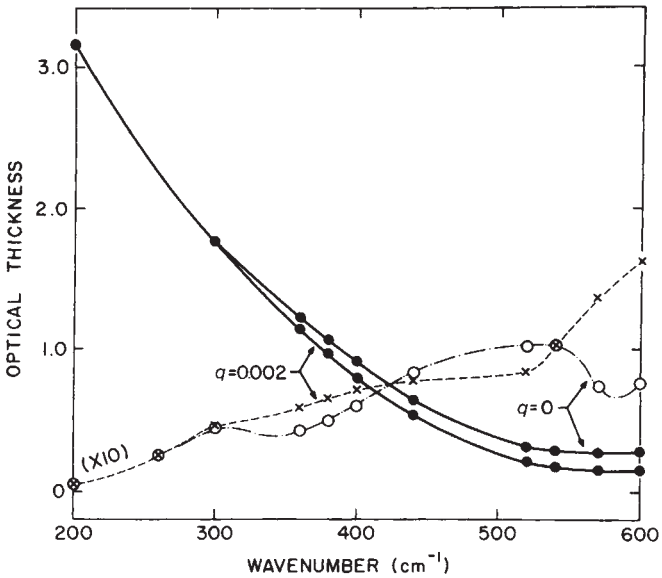


Fig. 6 Modified optical thicknesses. Smooth fits to $\Delta\tau_T$ in Fig. 5*a* for two values of q are shown (—●—). Corresponding solutions for $\Delta\tau_S$ are obtained by forcing a fit to the limb spectrum in Fig. 1 (— — × — — ×, and — · —○— · —○). The surface temperature is $T_G = 95$ K.

occultation T/M value of 3.32 K (AMU)$^{-1}$ for the surface[11] it becomes straightforward to calculate the surface temperature for different mole fractions of methane. As the tropopause temperature of 74 K is an upper limit, the surface temperatures scaled from this value are also upper limits.

The results from these calculations are shown in Fig. 4 as a function of the fraction of methane saturation. Also shown is the lower limit to the surface temperature calculated previously. Both limits are displayed as bands to illustrate the attendant uncertainties. Any combination of surface temperature and fraction of saturation within the enclosed triangle is a legitimate solution. Values for the fraction of saturation below 0.19 are not very likely on physical grounds, however, because they correspond to methane mole fractions less than the tropopause value. The lower troposphere seems to be undersaturated with respect to methane, perhaps substantially so, although the fairly long chain of reasoning involved makes it difficult to assess the associated uncertainties. From Fig. 4, with the fraction of saturation equal to 0.19, the limits for the surface temperature are

$$94\,K < T_G < 97\text{ K} \tag{6}$$

from which it follows that the limits for the mean molecular weight are

$$28.3 < M < 29.2\text{ AMU} \tag{7}$$

If the CH_4 fraction of saturation at the surface is >0.19, as is likely, the upper limits in expressions (6) and (7) must be reduced accordingly.

Tyler and Lindal (personal communication) have pointed out that the radio occultation values for T/M (ref. 11) are accurate to two significant figures but probably not to three, adding an uncertainty of about ±0.8 AMU to the limits of M in equation (7). On the other hand, inferring the surface temperature from one that is measured at the tropopause depends essentially only on the ratio of the respective T/M values. As this ratio is better determined than the individual values of T/M (G. L. Tyler and G. Lindal, personal communications), the upper limits for the surface temperature in Fig. 4 and equation (6) are well established.

Despite the many uncertainties M is probably >28, the value for molecular nitrogen. Taking $T_G = 95$ K and $(T/M)_G = 3.32$, we find $M = 28.6$. If $q(CH_4) = 0.06$ at the surface (<50% of saturation; see Fig. 4) and the atmosphere is composed only of hydrogen, methane, nitrogen and argon, it is possible to construct the mole fraction for each. The results are listed in Table 1 and show a large amount of argon (~12%), although the atmosphere remains dominantly nitrogen. Table 1 is only meant to be representative of reasonable conditions. We have no direct evidence that argon exists in the atmosphere, only that some molecular constituent higher in mass than N_2 is probably present, and argon is a plausible choice because of its inert qualities and low condensation temperature. The upper limit of 6% argon inferred by the UV experiment[15] is not inconsistent with our suggested value.

Molecular hydrogen abundance

Superimposed on the general tropopause opacity continuum shown in Fig. 3 are two localized broad features at 360 cm^{-1} and 600 cm^{-1} which correspond closely to the $S(0)$ and $S(1)$ pressure-induced absorption lines of molecular hydrogen. We have calculated approximate absorption coefficients for the pressure-induced absorption of H_2 due to collisions with N_2. The calculations are based primarily on the theoretical work of Birnbaum and Cohen[16] for H_2–H_2 collisions. The conversion to absorption coefficients for H_2–N_2 collisions was accomplished by scaling the laboratory data of Kiss and Welsh[17] and Kiss *et al.*[18]. As both the integrated absorption coefficient and the line half widths were scaled from data obtained at 195 K, there is an estimated uncertainty in the absorption coefficients of ~25% at the temperatures prevailing near the surface of Titan.

We have normalized the temperature profile in Fig. 2 to a surface temperature of 95 K, a realistic value according to Fig. 4. Induced H_2–N_2 absorption is then added to the other sources of opacity in the radiative transfer calculations. This new opacity source has a pressure-squared dependence, most dominant in the lower troposphere. Above the tropopause the effect was neglected because it is <1% of its value near the surface.

According to the models discussed previously, the tropopause opacity calculations are basically independent of the stratospheric model adopted (see Fig. 3). This remains true when pressure-induced H_2 opacity is added to the troposphere. Figure 5*a* shows several solutions for the optical thicknesses of the stratosphere and tropopause for the thin haze model with $T_S = 103.3$ K. Other stratospheric models yield essentially the same results for $\Delta\tau_T$. Thus the H_2 mole fraction seems fairly well defined if one imposes the condition that the wavenumber dependence of $\Delta\tau_T$ be as smooth as possible. A hydrogen mole fraction $q \sim 0.002 \pm 0.0005$ provides the smoothest fit to $\Delta\tau_T$ of the cases examined, although the actual uncertainty in q is larger owing to the 25% uncertainty in the hydrogen absorption coefficients, leading to a final value of

$$q(H_2) = 0.002 \pm 0.001 \tag{8}$$

The result for q is basically independent of the stratospheric model adopted. Further tests show this to be true for the tropospheric opacity also. One such test, shown in Fig. 5*b*, entails a model that departs radically from the previous ones. Instead of placing all the tropospheric opacity at the tropopause, substantial pressure-induced absorption associated with a fictitious gas is introduced, which places most of the absorption near the surface. The total optical depth of this absorber is shown as a function of wavenumber in Fig. 5*b*. The much weaker pressure-induced absorption of H_2 with $q = 0.002$ is then added, and solutions for $\Delta\tau_S$ and $\Delta\tau_T$ are obtained as before. Although the quantitative solutions for $\Delta\tau_T$ are substantially different from those in Fig. 5*a*, the shapes of the two curves corresponding to identical values of q are quite similar, showing the same degrees of smoothness. The solutions for $\Delta\tau_S$ are nearly identical in both cases. Thus the determination of q seems to be independent of the method by which the tropospheric background opacity is modelled as long as the wavenumber dependence of the latter is smooth.

An estimate of the sensitivity of the solutions to the accuracy of the data can be obtained by replacing $\Delta\tau_T$ with a completely smooth curve and forcing perfect fits to the limb radiances I_ν (52.7°) (see Fig. 1). Subjectively determined smooth fits of $\Delta\tau_T$ from Fig. 5 for $q = 0$ and $q = 0.002$ are shown in Fig. 6. Solutions for $\Delta\tau_S$ for $q = 0.002$ and 0 are also shown. These results lead in

© 1981 Nature Publishing Group

turn to the solutions for I_ν (7.3°) shown in Fig. 1 for $q = 0$ and $q = 0.002$; the observed spectrum at $\theta = 7.3°$ is superimposed.

As Fig. 1 demonstrates, the solutions for $q = 0.002$ fit the data well, whereas those for $q = 0$ show marked departures across the $S(0)$ line and smaller departures in the vicinity of the $S(1)$ line. In Fig. 6 the $q = 0$ solutions for $\Delta\tau_S$ inversely mimic the signatures corresponding in position to the $S(0)$ and $S(1)$ lines. Apparently this is how the model compensates for a lack of H_2 absorption in the troposphere in the process of fitting the limb spectrum in Fig. 1 (see also equation (5)). Conversely the $q = 0.002$ solutions for $\Delta\tau_S$ in Fig. 6 show no such inclination, implying that the H_2 opacity has been modelled correctly.

The total depth of our model atmosphere is ~90 km amagat, yielding from equation (8) a molecular hydrogen abundance of ~0.2 km amagat. This is a factor 25 lower than Trafton's[19,20] suggested value of 5 ± 3 km amagat, and a factor 10 lower than his lower limit. We cannot reconcile the two results. However, our value is consistent with the upper limit of 1 km amagat reported by Münch *et al.*[21]. It is also compatible with the steady-state mole fraction expected from balancing the production rate of H_2 (due to photochemical and charged particle decomposition of methane) with the H_2 loss rate (due to upward diffusion and subsequent blowoff at the top of the exosphere)[6].

We thank J. Pearl for helpful discussions and L. Horn and R. Koppany for observational planning; also G. L. Tyler and G. Lindal for a numerical listing for the radio occultation ingress profile, and R. Courtin for pressure-induced absorption calculations for various CH_4–N_2 mixtures in advance of publication.

Received 13 May; accepted 20 July 1981.

1. Gillett, F. C., Forrest, W. J. & Merrill, H. M. *Astrophys. J. Lett.* **184**, L93 (1973).
2. Low, F. J. & Rieke, G. H. *Astrophys. J. Lett.* **190**, L143–L145 (1974).
3. Gillett, F. C. *Astrophys. J. Lett.* **201**, L41–L43 (1975).
4. Danielson, R. E., Caldwell, J. J. & Larach, D. R. *Icarus* **20**, 437 (1973).
5. Caldwell, J. *Planetary Satellites* (ed. Burns, J.) 438 (University of Arizona Press, 1977).
6. Hunten, D. M. *NASA Conf. Publ. 2068* 127 (1978).
7. McCarthy, J. F., Pollack, J. B., Houck, J. R. & Forrest, W. J. *Astrophys. J.* **263**, 201–205 (1980).
8. Hanel, R. *et al.* Science **212**, 192–200 (1981).
9. Flasar, F. M., Samuelson, R. E. & Conrath, B. J. *Nature* **292**, 693–698 (1981).
10. Smith, B. A. *et al. Science* **212**, 163–191 (1981).
11. Tyler, G. L. *et al. Science* **212**, 201–206 (1981).
12. Savoie, R. & Fournier, R. P. *Chem. phys. Lett.* **7**, 1 (1970).
13. Obriot, J., Fondère, F., Marteau, P., Vu, H. & Kobashi, K. *Chem. phys. Lett.* **60**, 90 (1978).
14. Prydz, R. & Goodwin, R. D. *J. chem. Thermodyn.* **4**, 127–133 (1972).
15. Broadfoot, A. L. *et al. Science* **212**, 206–211 (1981).
16. Birnbaum, G. & Cohen, E. R. *Can. J. Phys.* **54**, 593–602 (1976).
17. Kiss, Z. J. & Welsh, H. L. *Can. J. Phys.* **37**, 1249–1259 (1959).
18. Kiss, Z. J., Gush, H. P. & Welsh, H. L. *Can. J. Phys.* **37**, 362–376 (1959).
19. Trafton, L. *Astrophys. J.* **175**, 285–293 (1972).
20. Trafton, L. *Icarus* **24**, 443–453 (1975).
21. Münch, G., Trauger, J. T. & Roesler, F. L. *Astrophys. J.* **216**, 963–966 (1977).

Titan's atmosphere: temperature and dynamics

F. M. Flasar, R. E. Samuelson & B. J. Conrath

Laboratory for Extraterrestrial Physics, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

In the lower atmosphere of Titan IR brightness temperatures exhibit meridional contrast ≲3 K. Seasonal variations are absent because of the large radiative time constant. In the upper stratosphere meridional contrasts are ~20 K, consistent with 100 m s^{-1} cyclostrophic zonal winds, and the radiative time constant is short, implying a large seasonal variation in the temperature and wind field. The absence of longitudinal thermal structure implies that zonally symmetric flows effect the meridional transport of heat. A simple model yields meridional velocities ~0.04 cm s^{-1} and vertical eddy viscosities ~10^3 cm^2 s^{-1} in the lower troposphere, and meridional velocities ~5 cm s^{-1} in the upper stratosphere.

COVERAGE of Titan by the IR instrument (IRIS) during the Voyager 1 encounter provided information on the global thermal structure of the satellite and its atmosphere at relatively high spatial resolution. Here we present data obtained in three spectral intervals which sample the lower troposphere and surface, tropopause, and the upper stratosphere. We discuss the meridional and longitudinal thermal structure implied by the data. Finally, we estimate radiative time constants at various levels of Titan's atmosphere and examine a class of dynamical flows which are consistent with the data.

Observations

The data were obtained over a 6-h interval when the spacecraft was within 2×10^5 km of the center of Titan and the instrument field of view was less than one-third of the satellite's 2,570 km radius. The season on Titan was close to northern spring equinox. Three spectral regions have been analysed. In the first, centred at 530 cm^{-1}, the total optical thickness of the atmosphere is $\tau \leq 0.6$ (ref. 1). The major contributor to the opacity is pressure-induced absorption of molecular hydrogen from H_2–N_2 collisions. As this effect is confined to a vertical range close to the ground, the 530 cm^{-1} brightness temperature should be close to the actual surface temperature. Brightness temperatures near 200 cm^{-1} correspond approximately to the tropopause temperature[1]. The 200 cm^{-1} opacity has been inferred to be fairly large ($\tau \geq 3$), confined mainly to clouds, and constrained to a vertical range over which the atmosphere is essentially isothermal. Finally, the 1,304 cm^{-1} ν_4 Q-branch of methane is associated with emission from the ~0.3 mbar pressure level.

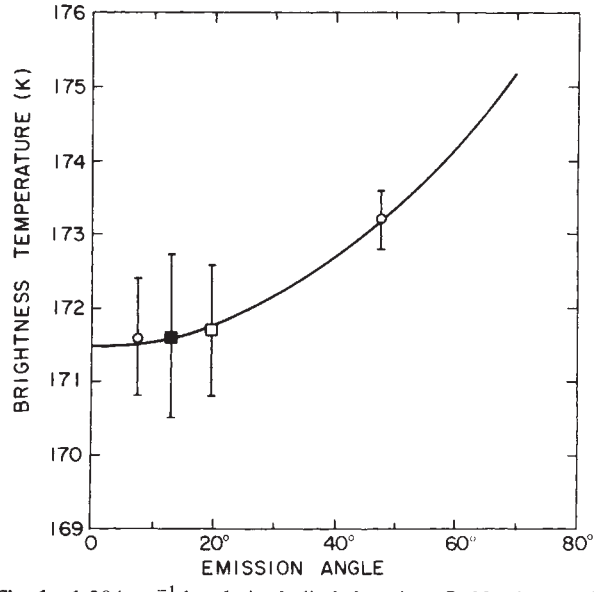


Fig. 1 1,304 cm^{-1} low latitude limb function. ○, Northern latitude daytime data; □, southern daytime data; ■, southern nighttime data. Error bars include uncertainties due to noise and calibration.

0028–0836/81/340693—06$01.00

© 1981 Macmillan Journals Ltd

Table 1 1,304–1,260 cm^{-1} brightness temperature differences

Latitude	Emission angle	$T_B(1,304)$	$T_B(1,260)$	δT_B(K)
−55.0°	50.5°	169.8±0.7	153.7±1.7	16.1±1.9
+7.5°	47.5°	173.2±0.4	156.7±1.3	16.5±1.4
+52.6°	50.5°	168.1±0.9	149.1±2.1	19.0±2.3

Vertical temperature gradients produce an emission angle dependence in the observed radiance, which must be removed to infer the horizontal structure in atmospheric temperature. We have therefore normalized the data in each wavenumber region to a common emission angle. Corrections at 200 cm^{-1} and 530 cm^{-1} were determined from a spectral limb function derived from data[1] restricted to the latitude range 0–15° N. These corrections have been applied to the data at all latitudes, even though there seems to be a subtle but real meridional change in opacity across the equator. There is no rigorous justification for doing this; however, as the corrections tend to be small (≤0.8 K at 200 cm^{-1} and ≤0.3 K at 530 cm^{-1}) any corresponding systematic error should also be small.

The larger emission angle corrections at 1,304 cm^{-1} are more reliable. Horizontal opacity inhomogeneities (due to clouds, for example) are not as likely near the 0.3 mbar level as they are at lower altitudes. Moreover, the difference between the brightness temperatures at 1,260 cm^{-1} and 1,304 cm^{-1}, divided by the difference in altitudes at which the respective CH_4 weighting functions peak, gives a crude approximation of dT/dz at each geographical location. Figure 1 depicts the emission angle corrections derived for low latitudes. Table 1 shows the spectral brightness temperature differences $\delta T_B = T_B(1,304) - T_B(1,260)$ at three widely separate latitudes where enough spectra are available for sufficient signal at 1,260 cm^{-1}. The emission angles are very similar, and the values of δT_B are identical to within the noise of the data. As CH_4 should be uniformly mixed at these levels, the altitude difference between weighting functions should be independent of geographical location. Thus the emission angle corrections derived for low latitudes should also be valid at high latitudes.

Other uncertainties and corrections exist. The statistical uncertainty due to noise was generally about ±1 K at 1,304 cm^{-1}; at 200 cm^{-1} and 530 cm^{-1} it was smaller, ±0.1 K, because it was possible to average over fairly wide spectral intervals (60–80 cm^{-1}) and because the IRIS signal-to-noise ratio was higher at these wavenumbers. Finally, all nightside data were corrected for a time-dependent calibration drift of the instrument. This drift occurred principally from the heating of the secondary mirror of the instrument which occurred when it became exposed to direct sunlight. The uncertainty in these calibration corrections is <0.1 K at 1,304 cm^{-1} and ~0.4 K and 0.5 K at 200 cm^{-1} and 530 cm^{-1}, respectively.

Results

The data have been examined for evidence of longitudinal structure. At several latitudes there were day/night observations separated by ~180° in longitude. At low latitudes longitudinal coverage was more extensive, though not complete; over limited spans of longitude, scales down to 30° could be resolved. In general fewer spectra could be averaged in this analysis than in the study of meridional structure, and the uncertainties exceeded the formal errors quoted above. Within these uncertainties, no longitudinal or hour angle dependence of brightness temperature was discernible. Reasonable upper limits on temperature variation are 1 K at 200 cm^{-1} and 530 cm^{-1} and 3 K at 1,304 cm^{-1}.

Meridional variations in temperature are better defined. Table 2 summarizes all the data used. The number of spectra used in obtaining the mean brightness temperatures T_B are listed and the mean latitude and emission angle α of the data set are given. Also indicated is the emission angle correction $\Delta T_B(\alpha)$ required to normalize T_B to $\alpha = 52.7°$. Finally, brightness temperatures corrected for emission angle are listed. The associated uncertainties are the sum of uncertainties due to noise, emission angle corrections, and nightside calibration corrections. Because the first represents a random error while the last two are systematic, we considered the sum of their magnitudes rather than a vector sum in estimating the total error. The corrected brightness temperatures are also shown in Fig. 2. Vertical bars represent the uncertainties listed in Table 2, while horizontal bars indicate the latitude range, defined by the centres of the relevant fields of view, over which the data set comprising each point extend.

The meridional variation in brightness temperature at 200 cm^{-1} is small. Brightness temperatures in the northern hemisphere are ~1 K higher than those at southern latitudes; a sharp transition occurs near 10° S. The meridional contrast in atmospheric temperature must also be small. The principal source of opacity at 200 cm^{-1} has been inferred to be a system of methane clouds, situated mainly at or just below the tropopause (~100 mbar) and ~3 in total optical thickness over northern low latitudes. No configuration of clouds with these opacities could mask meridional contrasts larger than a few degrees, as the temperature lapse rate is small near the tropopause. The small difference in brightness temperature between northern and southern latitudes may be an opacity rather than a thermal effect. As the abundance of stratospheric methane is controlled by saturation just below the tropopause[1,2], the methane cloud system will not extend very far into the stratosphere. An increase to $\tau \sim 4$ over southern latitudes would easily account for the ~1 K brightness temperature reduction. The correlation between our data and the enhanced visual brightness in the southern hemisphere found by Smith *et al.*[3] also suggests methane clouds. The opacity 'boundary' at ~10° S latitude seems to correlate fairly well with that discovered by Smith *et al.*[3] within about ±5° of the equator.

At 530 cm^{-1}, where the effective emission level is close to the ground[1], there is a real temperature variation with latitude. Pressure-induced H_2 absorption is the major opacity source while that from clouds is relatively small, implying that the effective emission level is perhaps ~1,500 mbar. The thermal variation with latitude is roughly symmetric about the equator. At low latitudes the daytime temperatures, which have smaller errors, are slightly warmer in the north than in the south. This may very well reflect a higher atmospheric opacity in the south, possibly from clouds at higher altitudes. This would be consistent with the north/south opacity difference inferred at 200 cm^{-1}. When extrapolated, the data suggest a total equator-to-pole surface temperature variation of about 3 K.

In contrast the meridional variation of the 1,304 cm^{-1}

Table 2 200, 530, and 1,304 cm^{-1} latitudinal temperature distributions

Day/night	No. spectra	Latitude (deg) 200 cm^{-1} 530 cm^{-1}	Latitude (deg) 1,304 cm^{-1}	Emission angle α(deg) 200 cm^{-1} 530 cm^{-1}	Emission angle α(deg) 1,304 cm^{-1}	T_B(K) 200 cm^{-1}	T_B(K) 530 cm^{-1}	T_B(K) 1,304 cm^{-1}	$\Delta T_B(\alpha)$ 200 cm^{-1}	$\Delta T_B(\alpha)$ 530 cm^{-1}	$\Delta T_B(\alpha)$ 1,304 cm^{-1}	T_B(corr) 200 cm^{-1}	T_B(corr) 530 cm^{-1}	T_B(corr) 1,304 cm^{-1}
N	13	−59.6	−55.0	54.6	50.5	72.8	91.3	169.8	+0.2	0	−0.2	73.0±0.5	91.3±0.6	169.6±0.7
N	8	−48.3	−45.0	45.8	42.0	73.3	91.3	173.5	−0.1	−0.1	+0.4	73.2±0.5	91.2±0.6	173.9±0.8
D	4	−29.6	−27.5	47.1	43.5	73.1	92.2	174.6	−0.5±0.2	−0.1	+0.4	72.6±0.3	92.1±0.2	175.0±1.0
D	6	−10.0	−9.5	21.0	19.5	74.1	92.8	171.7	−0.6±0.1	−0.3	+1.4	73.5±0.2	92.5±0.1	173.1±0.9
N	5	−8.7	−8.7	13.7	12.8	74.8	92.6	171.6	−0.7	−0.4	+1.6	74.1±0.5	92.2±0.6	173.2±1.1
D	29	+7.2	+7.5	52.7	47.5	74.0	93.1	173.2	0	0	0	74.0±0	93.1±0	173.2±0.4
D	10	+8.3	+8.4	7.3	7.3	74.8	93.5	171.6	−0.8	−0.4	+1.6	74.0±0	93.1±0.1	173.2±0.8
D	5	+33.9	+32.0	30.9	29.8	74.5	93.1	171.6	−0.5±0.2	−0.3	+1.0	74.0±0.2	92.8±0.1	172.6±1.1
N	6	+35.7	+33.5	45.8	42.0	74.2	91.9	171.2	−0.2±0.1	−0.1	+0.4	74.0±0.6	91.8±0.6	171.6±1.1
D	10	+57.1	+52.6	55.0	50.5	73.6	91.3	168.3	+0.1	0	−0.2	73.7±0.1	91.3±0.1	168.1±0.9
D	6	+67.4	+61.4	66.9	59.2	73.7	90.8	164.8	+0.7±0.4	+0.3±0.1	−1.0	74.4±0.5	91.1±0.3	163.8±1.4

© 1981 Nature Publishing Group

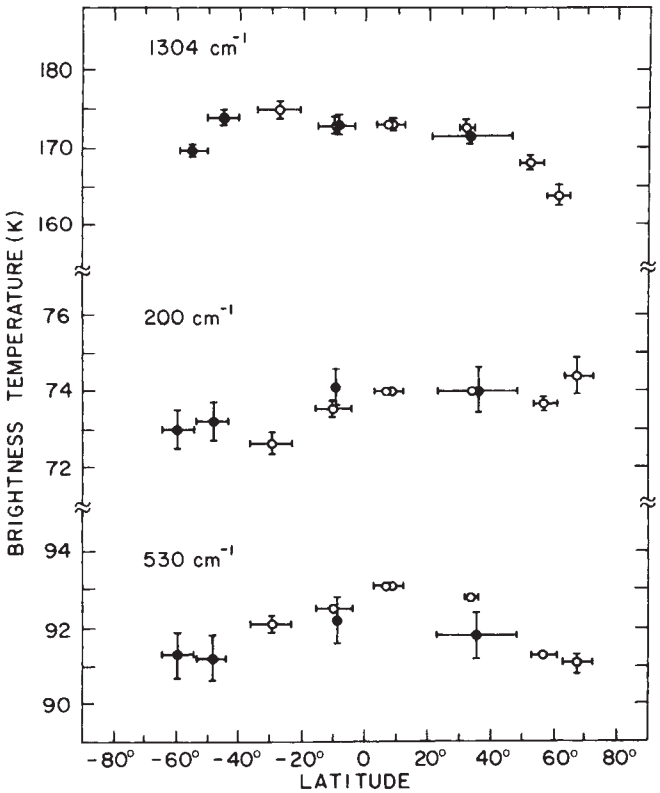


Fig. 2 1,304 cm^{-1}, 530 cm^{-1}, and 200 cm^{-1} latitudinal distributions of brightness temperature. The 1,304 cm^{-1} distribution is normalized to an emission angle $\alpha = 47.5°$, while the other two are normalised to $\alpha = 52.7°$. ○, Daytime data; ●, night-time data. Vertical bars include uncertainties due to noise, calibration, and emission angle corrections. Horizontal bars denote the latitude range over which data comprising each point extend.

brightness temperature is much larger. Although the extrapolation is large, the poles are apparently 20 K colder than the equatorial region. There is a hemispheric asymmetry, the southern hemisphere being 3 K warmer at mid latitudes than the northern hemisphere. The structure is probably thermal in origin with little or no contribution from meridional opacity variations along constant pressure surfaces. After correcting for an emission angle of 47.5° (to which the data are normalized) the effective emission angle is less than that for the lower wavenumbers ($\alpha = 52.7°$), because of its functional dependence on altitude in an extended spherical atmosphere. This also accounts for the slight compression of the 1,304 cm^{-1} data along the latitude axis relative to the data at the lower wavenumbers.

Discussion

The observed thermal structure of an atmosphere represents the radiative–dynamical response to an imposed thermal driving, such as solar heating. An important measure of the inertia of an atmosphere to this driving is its radiative relaxation, or response time τ_R, (ref. 4). Smith *et al.*[3] have estimated the radiative response time of Titan's lower troposphere to be $\sim 4 \times 10^9$ s, or 140 yr. This is much longer than a season on Titan. If $\Omega_s = 2\pi/$(Saturn's orbital period $\simeq 29.5$ yr), $\Omega_s \tau_R \simeq 30$. Such a large value implies almost no seasonal variation in temperature. At each latitude, temperatures should respond to the annual mean insolation which is symmetric about the equator. The symmetry evident in the brightness temperatures at 530 cm^{-1} is consistent with this.

Estimates of the radiative response time near the tropopause, to which the 200 cm^{-1} brightness temperatures in Fig. 2 pertain, are not presently feasible. Most of the tropopause emission is at wavenumbers below 200 cm^{-1} for which we have no data. The low meridional contrast in brightness temperature at 200 cm^{-1} suggests little seasonal variation in temperature. Either the radiative time constant at the tropopause is large, $\Omega_s \tau_R \gg 1$, or the heating at this level is predominantly from absorption of radiation emitted lower in the atmosphere where τ_R is large and seasonal effects are small.

Radiative time constants can be estimated for Titan's stratosphere. The contribution from gaseous emission is taken into account using the formulation of Harshvardhan and Cess[5], which assumes direct cooling to space. The cooling rate N (s^{-1}) at a pressure level P is

$$N = -\frac{\pi g}{C_p P} \Delta\nu \frac{\partial \tau_{\Delta\nu}}{\partial \ln P} \frac{dB_{\nu_0}}{dT} \qquad (1)$$

where $\tau_{\Delta\nu}$ is the transmittance averaged over a wavenumber interval $\Delta\nu$ (cm^{-1}), B_{ν_0} is the Planck function (erg s^{-1} cm^{-1} sr^{-1}) at ν_0, g is the local gravitational acceleration, T is temperature, and C_p is the specific heat. Transmittances have been obtained from a line by line spectral integration for CH_4 ($\nu_0 = 1{,}304$ cm^{-1}, $\Delta\nu = 200$ cm^{-1}), C_2H_6 ($\nu_0 \simeq 820$ cm^{-1}, $\Delta\nu = 85$ cm^{-1}), and C_2H_2 ($\nu_0 \simeq 729$ cm^{-1}, $\Delta\nu = 95$ cm^{-1}) (V. Kunde personal communication). Each gas was assumed to be distributed uniformly with height in the stratosphere with an abundance inferred from its observed IR emission[2]. Radiative response times were obtained by summing over the three species (Table 3).

Radiative response times in the upper stratosphere (0.3 mbar $< P <$ 10 mbar) are short relative to Titan's seasons but still long compared with its day. Note that the times are smaller than the preliminary estimates given by Hanel *et al.*[2]. The longer radiative time constants in the lower stratosphere mainly result from the decrease in T (and hence dB_{ν_0}/dT) with depth in this region[2,6]. The assumption of a uniform vertical distribution is valid only if there is rapid vertical mixing. If C_2H_6 and C_2H_2 are end products of the photo-dissociation of CH_4 at high levels of the atmosphere ($P \ll 0.1$ mbar)[7], less efficient mixing would imply that the concentrations of C_2H_6 and C_2H_2 increase with height. This would effect a decrease in response times in the upper stratosphere, and an increase at lower levels, relative to those listed in Table 3.

It has been concluded[1] that some of the emission at 200–600 cm^{-1} originates in the stratosphere, probably from aerosols. As the stratosphere is optically thin in this wavenumber range, the level of emission cannot be unambiguously determined. However, the deduced wavelength dependence of the aerosol emissivity is more like that expected for small particles if the emission is from the lower stratosphere ($P \sim 10$–30 mbar). If the aerosol emission is confined to a slab one scale height thick with emissivity ε_ν and temperature T, the radiative response time of the slab is given by

$$\tau_R^{-1} = \frac{2\pi g}{C_p P} \int_{200\ cm^{-1}}^{600\ cm^{-1}} \varepsilon_\nu \frac{dB_\nu}{dT} d\nu \qquad (2)$$

Table 3 lists τ_R at two pressure levels. The aerosol component reduces τ_R significantly below that computed for gaseous emission alone. This suggests that aerosols are the major source of cooling in the lower stratosphere.

In the upper stratosphere near 1 mbar $\Omega_s \tau_R \ll 1$, and there should be no lag in the thermal response. As only a small fraction of incident solar radiation is absorbed at higher altitudes, the meridional distribution of solar heating and temperature at solstice will be flat in the summer hemisphere, with a slight maximum at the pole, provided that the sources of thermal and solar opacity are uniformly distributed with latitude[8,9]. The winter hemisphere will be colder with sharper gradients in temperature. At the equinoxes the temperature distribution

Table 3 Radiative response times in Titan's stratosphere

		Gaseous radiative cooling rate N(s^{-1})			τ_R(s)		$\Omega_s \tau_R$	
P(mbar)	T(K)	CH_4	C_2H_6	C_2H_2	Gas	Aerosol	Gas	Aerosol
0.3	173	2×10^{-8}	3×10^{-8}	2×10^{-8}	2×10^7	—	0.1	—
1	170	1×10^{-8}	7×10^{-9}	1×10^{-8}	3×10^7	—	0.2	—
10	149	1×10^{-9}	4×10^{-9}	4×10^{-9}	1×10^8	4×10^7	0.7	0.2
30	104	3×10^{-12}	2×10^{-10}	5×10^{-10}	1×10^9	7×10^7	9.0	0.5

© 1981 Nature Publishing Group

should be symmetric about the equator, being warmest there and coldest at the poles. The meridional distribution of the 1,304 cm^{-1} brightness temperature at northern spring equinox (Fig. 2) roughly accords with the latter case, but the ~3 K hemispheric asymmetry evident at mid-latitudes is puzzling. Meridional variations in the infrared flux emitted deeper in the atmosphere where $\Omega_s\tau_R \gtrsim 1$ cannot explain the asymmetry, as the flux is emitted at lower temperatures and does not contribute significantly to the radiative heating near 1 mbar. A hemispheric asymmetry in opacity is possible. On Earth, for example, seasonal variations in the meridional distribution of ozone account for the seasonal behaviour of stratospheric temperatures[10]. Although variations in aerosols and gaseous absorbers such as C_2H_2 and C_2H_6 are possible on Titan, we cannot unambiguously account for the observed asymmetry.

Figure 3 represents schematically $\Omega_s\tau_R$ with pressure level. Because of the uncertainties in the location of the stratospheric aerosol emission and in the radiative time constant at the tropopause, the transition from $\Omega_s\tau_R \ll 1$ high in the stratosphere to $\Omega_s\tau_R \gg 1$ lower down is not defined very precisely.

The meridional temperature gradients depicted in Fig. 2 imply a zonal thermal wind, u. If the surface drag is sufficiently strong, $u(z=0)\approx 0$, and

$$\frac{u^2(z)}{a}\tan\Lambda + 2\Omega\sin\Lambda\, u(z) = -\frac{1}{a}\int_0^z \frac{g}{T}\partial_\Lambda T\, dz \qquad (3)$$

where Λ is latitude, z is height, $a = 2{,}570$ km is the radius of Titan[6], $\Omega = 4.5\times10^6\ s^{-1}$ is its frequency of rotation, and T is atmospheric temperature. If we assume that the meridional temperature has the form $T(z,\Lambda) = T_0(z) + \Delta T(z)\cos\Lambda$ and $\Delta T \approx 3$ K throughout the troposphere, then at the tropopause ($z \approx 60$ km, $P \approx 100$ mbar) we infer an eastward velocity of $u \approx 36\ m\ s^{-1}$ at 45° latitude. This is approximately three times the surface rotational speed at Titan's equator. The larger horizontal contrasts in temperature in the stratosphere imply even stronger zonal winds. We assume that the thermal contrast between the equator and pole increases linearly with $\ln P$ from 3 K at 100 mbar to 20 K at 0.5 mbar. At 45°, equation (3) yields $u(0.5\ \text{mbar}) \approx 110\ m\ s^{-1}$. The estimate of the stratospheric zonal winds is uncertain insofar as we have no direct knowledge of the meridional temperature structure in the lower stratosphere. Nonetheless, it is hard to avoid the conclusion that there are global cyclostrophic flows ($u/\Omega a \gg 1$) in Titan's atmosphere. The occurrence of such flows in the atmosphere of Venus is already well established[11]. However, unlike Venus, Titan has a large obliquity. The temperatures in its upper stratosphere, where $\Omega_s\tau_R$ is small compared with one, should exhibit a large seasonal variation. From equation (3), we expect a concomitant variation in the zonal winds at high altitudes.

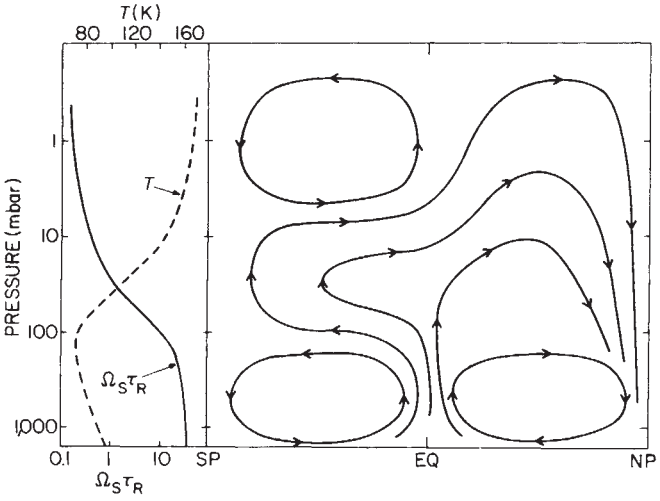


Fig. 3 Schematic diagram of the meridional flow in Titan's atmosphere at northern spring equinox. Mean vertical profiles of temperature and radiative relaxation time are indicated at the left.

Table 4 Radiative and dynamical characteristics of Titan's atmosphere

	$\Delta\theta$ (K)	D (km)	$\delta\theta$ (K)	τ_R (s)	V ($cm\ s^{-1}$)	a/V (s)	κ ($cm^2\ s^{-1}$)
Lower troposphere ($P \sim 1{,}000$ mbar)	3	28	22	4×10^9	0.04	7×10^9	10^3
Upper stratosphere ($P \sim 1$ mbar)	20	42	60	3×10^7	5	6×10^7	—

The condition that $u(0)$ be small compared with the zonal winds at higher levels requires that the time for surface drag to deplete the relative zonal momentum of the atmosphere be small compared with the other dynamical time scales of the problem[12]. We assume a surface stress of the form $C_D u(0)^2$, where $C_D \approx 3\times10^{-3}$ is a dimensionless drag coefficient[13]. For vertical scales ($D \approx 28$ km) and zonal velocities characteristic of the lower troposphere, the surface drag time is only[12] $D/C_D u(D) \sim 10^6$ s. We will show that the meridional turnover time is much larger, $\sim 10^{10}$ s (Table 4).

The long radiative time scales in Titan's atmosphere imply that there should be no diurnal response to solar heating and this is consistent with the absence of discernible longitudinal structure in the data. The absence of this structure, however, also implies that baroclinic eddies do not have an important role in Titan's atmosphere. Previously, Leovy and Pollack[13] concluded that baroclinic eddies are probably not effective in transporting heat on Titan. Assuming a vertical structure close to adiabatic, they found that the preferred scale of the eddies exceeded the radius of the satellite. Voyager observations, however, indicate a temperature lapse rate which is much more stable[2,6]. This increases the preferred scale of the eddies and strengthens Leovy and Pollack's original conclusion that zonally symmetric flow constitutes the favoured mode of meridional heat transport.

Following Leovy and Pollack[13], we consider zonally symmetric, thermodynamically direct, meridional circulations of horizontal dimension $\sim a$. There are several reasons for considering only thermally direct flows. First, the thermal contrasts forced by differential solar heating will induce a meridional transport of heat which will reduce these contrasts for the atmosphere as a whole. Second, the observed temperatures in the lower troposphere and upper stratosphere, the two regions of the atmosphere which we can treat quantitatively, are consistent with thermally direct flows. The global temperature contrasts are in the same sense as those which would obtain from purely radiative effects but reduced in magnitude. Finally, a thermally direct circulation is the simplest, and allows straightforward calculation, provided that the meridional contrasts in temperature and radiative time constants are known. Given the gaps in our knowledge of Titan's atmosphere, consideration of more complex flows is unwarranted.

We use here the Boussinesq approximation, although this limitation does not affect our conclusions. Consider an idealized meridional cell confined to altitudes z_0 and $z_0 + D$, and latitudes 0 and $\pi/2$. Let the symbol $\delta(\ )$ denote the vertical contrast in a given quantity ($\delta(\ ) = (\ )_{z=z_0+D} - (\ )_{z=z_0}$) averaged over latitudes 0, $\pi/2$; Let $\Delta(\ )$ denote the contrast over latitude ($\Delta(\ ) = (\ )_{\Lambda=0} - (\ )_{\Lambda=\pi/2}$) averaged over the depth D. The thermodynamic heat equation can be cast into the form[24]:

$$\partial_t\Delta\theta + \frac{V}{a}\delta\theta \sim \frac{\Delta\theta_e - \Delta\theta}{\tau_R} \qquad (4)$$

where θ denotes potential temperature, and factors of the order of unity have been neglected. The term on the right-hand side represents radiative heating and cooling in terms of a linear relaxation back to the radiative equilibrium state, θ_e. The second term on the left-hand side represents the poleward transport of sensible heat and potential energy by the meridional cell. V is a mass-weighted meridional velocity averaged over the upper or lower leg of the cell. When $\Delta\theta > 0$, there is upwelling at low latitudes and poleward motion in the upper leg of the cell. The condition $\delta\theta > 0$ ensures that net transport of energy is poleward[14], and the circulation acts to reduce $\Delta\theta$ ($\partial_t\Delta\theta < 0$). We will use equation (4) to estimate the magnitude of the meridional circulation at various altitudes in Titan's atmosphere.

© 1981 Nature Publishing Group

In the troposphere ($z_0 = 0$) the long radiative response time and absence of detectable seasonal variation imply a steady circulation, $\partial_t \equiv 0$, symmetric about the equator. As the lower troposphere contains most of the atmospheric mass, we choose D = 1.5 scale heights ≈28 km. Over this interval $\delta\theta \approx 22$ K (refs 2, 4). As the inclination of Titan's rotational axis to the ecliptic is presumably 27°, the annual mean insolation of the polar regions is half that at equatorial latitudes[15]. A crude estimate of $\Delta\theta_e$ follows from

$$\left(\frac{\theta_e(\Lambda=0)}{\theta_e\left(\Lambda=\frac{\pi}{2}\right)}\right)^4 \sim 2.$$

With $\theta_e(\Lambda=0) \approx 95$ K, $\Delta\theta_e \approx 15$ K. The temperature contrast observed in the troposphere (Fig. 2) is much less, $\Delta\theta \sim 3$ K. With these estimates, equation (4) implies $V \sim 0.04\ \mathrm{cm\,s^{-1}}$, equivalent to a dynamical turnover time, a/V, $\sim 7\times10^9$ s ~200 yr (Table 4).

Two potentially important transports have been neglected in this discussion of the Hadley circulation. The first is the equatorward transport by the circulation of latent heat associated with methane condensation. The second is the poleward transport of sensible heat by deep reservoirs of liquid methane which may exist on the surface of Titan. The magnitude of these transports relative to the atmospheric transport of sensible heat and potential energy is difficult to estimate. On Earth, where water is the condensate, they are all comparable, at least over those latitudes where the Hadley circulation dominates[16]. In this complex situation, the net poleward transport of energy on Earth is still comparable to the atmospheric transport of sensible heat and potential energy alone. This suggests that our estimates remain correct to order of magnitude.

In the upper stratosphere the radiative response times are short. Since $\partial_t \sim \Omega_S \ll \tau_R^{-1}$, the time derivative in equation (4) can be neglected: the upper stratosphere is in radiative–dynamical balance with the current value of the seasonal forcing. For the estimate we assume a vertical scale of one scale height. At the equinoxes, radiative equilibrium implies $\Delta\theta_e \sim 10^2$ K. Using the thermal contrast $\Delta\theta \sim 20$ K inferred from the 1,304 $\mathrm{cm^{-1}}$ data, we estimate $V \sim 5\ \mathrm{cm\,s^{-1}}$ (Table 4).

The transition from $\Omega_s\tau_R \ll 1$ in the upper stratosphere to $\Omega_s\tau_R \gg 1$ occurs in the lower stratosphere and tropopause region, as noted earlier. In the lower stratosphere the thermal response will lag the solar forcing; if $\Omega_s\tau_R \sim 3$ the lag is close to its asymptotic value of one season[17]. At the northern spring equinox, temperatures within the lower stratosphere should be warmer in the southern hemisphere than in the north. Such a thermal asymmetry would induce a pole to pole circulation with rising motion in the southern hemisphere. If such a circulation penetrated into the tropopause region, where the seasonally varying radiative heating is probably smaller, adiabatic cooling could conceivably be sufficiently strong to produce lower temperatures and enhanced condensation in the southern hemisphere, as suggested by the observations at 200 $\mathrm{cm^{-1}}$. On the basis of the visible hemispheric asymmetry, Smith *et al.*[3] suggested such a circulation for the lower troposphere. However, the magnitudes of radiative response times (Tables 3 and 4) and the global hemispheric symmetry observed at 530 $\mathrm{cm^{-1}}$ (Fig. 2) suggest that such a cross equatorial circulation is more likely situated in the lower stratosphere and tropopause region.

Figure 3 illustrates one possible model of the meridional flow in Titan's troposphere and stratosphere at northern spring equinox; it assumes a uniform distribution of opacity with latitude. The topology indicated has to be speculative, because we lack sufficient information on the vertical structure of the meridional contrasts in temperature. This cannot be achieved until the vertical distribution of IR opacity is known, particularly in the lower stratosphere and troposphere. Note, however, that the vertical mass flux in the ascending leg of the meridional cell in the upper stratosphere is 0.1 that in the lower troposphere. Thus the streamlines in the stratospheric circulation possibly close in the lower troposphere, as apparently occurs on Earth[18,19].

Although eddy transport of heat may not be important on Titan, eddies are crucial in the maintenance of its zonal momentum. The zonal velocity of 36 $\mathrm{m\,s^{-1}}$ at 45° latitude, computed earlier, corresponds to an angular momentum per unit mass of $3\Omega a^2$. In the upper stratosphere the corresponding angular momentum per unit mass is $2\frac{1}{2}$ times larger. The angular momentum per unit mass in a zonally symmetric circulation controlled by down-gradient angular momentum fluxes is limited to Ωa^2 (refs 12, 20). Atmospheres with excess angular momentum therefore require other transport mechanisms, such as lateral mixing of vorticity or angular velocity by barotropic eddies[21–23]. Upward transport of zonal momentum by vertically propagating waves is another possibility[23], although the lack of longitudinal structure in the observations precludes most conceivable planetary-scale waves on Titan.

The vertical transport of angular momentum by a zonally symmetric meridional cell must be balanced in a steady state by vertical eddy transport. The angular momentum equation, averaged over a horizontal surface from $\Lambda = 0$ to $\pi/2$, is [21]

$$\partial_t \int_0^{\pi/2} \mathrm{d}\Lambda \cos\Lambda M + \int_0^{\pi/2} \mathrm{d}\Lambda \cos\Lambda \partial_z (WM) = \int_0^{\pi/2} \mathrm{d}\Lambda \cos\Lambda \partial_z(\kappa\partial_z M) \quad (5)$$

In equation (5) M denotes the angular momentum per unit mass, W is vertical velocity, and κ is an eddy diffusion coefficient for vertical mixing of momentum. We have assumed the horizontal fluxes to be zero at $\Lambda = 0$, $\pi/2$. Consider the lower troposphere which contains the bulk of the atmospheric mass and momentum; here the flow is steady ($\partial_t \equiv 0$). The requirement that there be no net exchange of angular momentum between the lower troposphere and surface implies that the vertical flux of angular momentum across a horizontal surface vanish, permitting equation (5) to be written as

$$W\Delta M \sim 2\frac{\kappa\,\delta M}{D} \quad (6)$$

The quantities ΔM and δM denote, respectively, the horizontal and vertical contrasts in angular momentum. An atmosphere with $\partial_z u > 0 (\delta M > 0)$ requires that the equatorial angular momentum exceed that at polar latitudes ($\Delta M > 0$)[21]. Knowing the meridional velocities from thermodynamic considerations (Table 4) permits an estimate of κ. Since $u(D) \sim \Omega a$ and $u(0) \approx 0$, in the lower atmosphere of Titan $\delta M \sim \langle M \rangle$, a tropospheric average. A Hadley cell which conserves angular momentum in its (upper) poleward branch will have $\Delta M \cong 0$. On the other hand, if eddy mixing of vorticity or angular velocity across latitudes is sufficiently vigorous, $\Delta M \sim \langle M \rangle$. Thus we expect $\Delta M/\delta M \lesssim 1$ and from equation (6)

$$\kappa \lesssim DW \sim \frac{D^2}{a} V \sim 10^3\ \mathrm{cm^2\,s^{-1}}$$

From analogous considerations, Gierasch[21] inferred a low value of vertical viscosity in the atmosphere of Venus, $\kappa \sim 10^4\ \mathrm{cm^{-1}}$. Note that equation (6), with the added condition $\Delta M/\delta M \lesssim 1$, requires that the time for diffusion of momentum vertically through the troposphere, D^2/κ, not exceed the dynamic turnover time, a/V.

Leovy and Pollack[13] proposed a symmetric model of Titan's tropospheric circulation and deduced meridional velocities much larger and thermal contrasts smaller than those presented here. Although their model differs from ours in several specific aspects, the most important difference lies in their treatment of internal friction. In not distinguishing between $u(0)$ and $u(D)$, they in effect assumed a Rayleigh friction drag in the angular momentum equation. They therefore implicitly assumed $\kappa \sim C_D u(D) D \sim 10^7\ \mathrm{cm^2\,s^{-1}}$, well in excess of our derived upper limit. From equation (6) this resulted in larger meridional velocities, $V \sim 1\ \mathrm{m\,s^{-1}}$, and smaller thermal contrasts. The

© 1981 Nature Publishing Group

anticipated seasonal variation in the zonal wind field of the stratosphere presents a more complex situation, and we will not attempt to analyse the requisite angular momentum transports.

Conclusions

IR brightness temperature at 200 cm^{-1}, 530 cm^{-1}, and 1,304 cm^{-1} exhibit small meridional contrasts ($\leqslant 3$ K) in the troposphere and tropopause region and larger contrasts (~ 20 K) in the upper stratosphere. Radiative relaxation times estimated for the troposphere and stratosphere are large compared with the length of Titan's day. In the lower troposphere they are also large compared with the length of a season, $\Omega_s \tau_R \gg 1$, implying a lack of seasonal variation in temperature. In the upper stratosphere, $\Omega_s \tau_R \ll 1$, and seasonal variations in temperature should be large. The location of the transition region, where $\Omega_s \tau_R \simeq 1$, is not well determined.

The absence of detectable longitudinal structure in the observations suggests that baroclinic waves and eddies are absent in Titan's atmosphere. Zonally symmetric flows are the preferred mode of meridional heat transport and thermally driven planetary-scale circulations yield characteristic meridional velocities of 0.04 cm s^{-1} in the lower troposphere and 5 cm s^{-1} in the upper stratosphere.

The observed meridional temperature contrasts imply zonal winds which are cyclostrophic, reaching ~ 100 m s^{-1} in the upper stratosphere. The anticipated seasonal variation in the stratospheric temperature field implies a concomitant variation in the zonal winds at high altitudes. The inferred cyclostrophic flow has an angular momentum per unit mass which, at least in the stratosphere, greatly exceeds Ωa^2. In the lower troposphere, an assumed balance in the vertical transport of angular momentum by zonally symmetric meridional flow and down gradient diffusion implies a vertical viscosity of only 10^3 cm^2 s^{-1}.

We thank V. G. Kunde for gaseous transmittance tables for Titan's stratosphere, W. C. Maguire for helpful discussions, also M. E. Geller, P. J. Gierasch, C. B. Leovy and R. A. Hanel for their critical reading of the manuscript.

Received 18 June; accepted 20 July 1981.

1. Samuelson, R. E., Hanel, R. A., Kunde, V. G. & Maguire, W. A. *Nature* **292**, 688–693 (1981).
2. Hanel, R. *et al. Science* **212**, 192–200 (1981).
3. Smith, B. A. *et al. Science* **212**, 162–190 (1981).
4. Gierasch, P. J., Goody, R. M. & Stone, P. H. *Geophys. Fluid Dyn.* **1**, 1–18 (1970).
5. Harshvardhan & Cess, R. D. *Tellus* **28**, 1–9 (1976).
6. Tyler, G. *et al. Science* **212**, 201–206 (1981).
7. Strobel, D. F. in *Atmospheres of Earth and the Planets* (ed. McCormac, B. M.) 401–408 (Reidel, Dordrecht, 1975).
8. Hess, S. L. *Introduction to Theoretical Meteorology* (Holt, Rinehart and Winston, New York, 1959).
9. Carlson, B. E., Caldwell, J. & Cess, R. D. *J. atmos. Sci.* **37**, 1883–1885 (1980).
10. Prabhakara, C., Rodgers, E. B., Conrath, B. J., Hanel, R. A. & Kunde, V. G. *J. geophys. Res.* **81**, 6391–6399 (1976).
11. Schubert, G. *et al. J. geophys. Res.* **85**, 8007–8025 (1980).
12. Held, I. M. & Hou, A. Y. *J. atmos. Sci.* **32**, 1038–1044 (1975).
13. Leovy, C. B. & Pollack, J. B. *Icarus* **19**, 195–201 (1973).
14. Stone, P. *J. atmos. Sci.* **32**, 1005–1016 (1975).
15. Brinkman, A. W. & McGregor, J. *Icarus* **38**, 479–482 (1979).
16. Lorenz, E. N. *The Nature and Theory of the General Circulation of the Atmosphere* World Meteorological Organization (1967).
17. Cess, R. D. & Caldwell, J. *Icarus* **38**, 349–357 (1979).
18. Geller, M. A. *J. atmos. terr. Phys.* **41**, 683–705 (1979).
19. Dunkerton, T. *J. atmos. Sci.* **35**, 2325–2333 (1978).
20. Schneider, E. K. *J. atmos. Sci.* **34**, 280–296 (1977).
21. Gierasch, P. J. *J. atmos. Sci.* **32**, 1038–1044 (1975).
22. Rossow, W. B. & Williams, G. P. *J. atmos. Sci.* **36**, 377–389 (1979).
23. Leovy, C. B. *J. atmos. Sci.* **30**, 1218–1220 (1973).
24. Stone, P. H. *J. atmos. Sci.* **31**, 1681–1690 (1974).

Implications of Titan's north–south brightness asymmetry

Lawrence A. Sromovsky*, Verner E. Suomi*, James B. Pollack†, Robert J. Krauss*, Sanjay S. Limaye*, Tobias Owen‡, Henry E. Revercomb* & Carl Sagan§

* Space Science and Engineering Center, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA
† NASA Ames Research Center, Moffett Field, California 94035, USA
‡ Department of Earth and Space Sciences, Suny at Stony Brook, Stony Brook, New York 11790, USA
§ Laboratory for Planetary Studies, Cornell University, Ithaca, New York 14853, USA

Voyager 1 images of Titan, when normalized to remove limb darkening, reveal an axially symmetric brightness pattern with significant north–south asymmetry. This interhemispheric contrast seems to be a response to seasonal solar heating variations resulting from Titan's inclined spin axis. The contrast significantly lags the solar forcing, indicating that its production involves the atmosphere well below the unit optical depth level. The contrast has a significant effect on Titan's disk-integrated brightness as seen from Earth, and probably accounts for most of the observed long term variation, with solar UV variations accounting for the remainder.

EARTH-based observations[1,2] show that between 1972 and 1976 Titan's disk-integrated brightness increased by ~9% in the blue (b) and ~5.5% in the yellow (y) to a maximum during 1976–77, and decreased through 1978. This variation and a similar but smaller variation in Neptune's brightness can be correlated with the solar cycle[2,3], suggesting solar variability as the common cause. As photochemical reactions are thought to produce the submicrometre aerosols in both atmospheres, we expect the variation in solar UV output during the solar cycle[4,5] to affect the amount or physical properties of the aerosols. Pollack *et al.*[6] showed that Titan's brightness changes during 1972–76 could be produced by changes in mean particle size, visible absorption coefficient, or optical depth, and that the fractional decrease in the particle production rate required to explain the increase in albedo was roughly consistent with the estimated fractional change in solar UV output.

A new mechanism for explaining at least part of the Earth based photometry of Titan is suggested by the Voyager 1 observations of November 1980 that showed Titan to be shrouded in an apparently unbroken cloud cover with several large-scale zonal features and a surprising north–south asymmetry in brightness[7]. As Titan's spin axis is inclined 27° to the ecliptic (assuming Titan is tidally locked to Saturn), the angle between Titan's axis and the line of sight to the Earth varies during Saturn's (and Titan's) 29.5-yr orbit around the Sun. The apparent tipping motion, combined with the brightness asymmetry could result in a long-term periodic variation of Titan's disk-integrated brightness. If the brightness asymmetry also varies in response to seasonal changes in solar heating of the deepest layers of the atmosphere[7], then the contrast would follow the seasonal heating asymmetry with a phase shift of nearly a full season and would reach an extreme value near Titan's equinoxes (near the time of the Voyager 1 encounter). Using a highly simplified model to combine the effects of the suggested seasonal changes in contrast and the time-dependent viewing conditions, the predicted brightness variation is found

0028–0836/81/340698—05$01.00

© 1981 Macmillan Journals Ltd

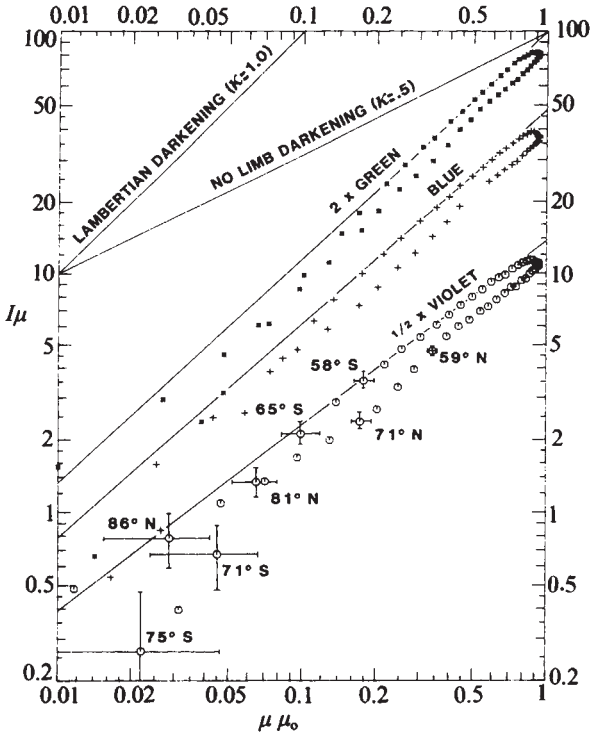


Fig. 1 Voyager 1 intensity measurements in violet, blue, and green filters, along Titan's 150° E meridian. Lines drawn for each colour show the Minnaert fits to intensities between 31° and 52° S. Intensities (in digital counts) have been corrected for dark noise but not converted to absolute units. The error bars labelled by latitude show the effects of an intensity uncertainty of ±0.5 count and an uncertainty in μ and μ_0 corresponding to a ±1 pixel error in the image location of Titan's centre. The identification of image line and element coordinates with Titan latitude and longitude, image navigation, is based on predicted spacecraft location and attitude information supplemented by a precise determination of camera pointing derived from the location of Titan's centre within the image.

to be in phase with the Earth-based photometry and within a factor of 2 of the observed magnitude change[7].

The essential differences between the solar cycle and seasonal components are the driving forces and the response times. The former is driven by the changing UV output of the Sun, and brightness changes in the same direction would be expected at all latitudes at the same time, with a delay of the order of six months relative to the UV forcing function[8]. The seasonal component is driven by the tipping of Titan's axis relative to the Sun, causing simultaneous brightness changes in different directions in northern and southern hemispheres, with a time lag of nearly a full season (~7 yr). In both cases, the properties of the aerosol particles are probably modulated and in turn produce the observed brightness changes.

We now use Voyager 1 and Pioneer 11 observations, with recent Earth-based photometry, to place firmer bounds on the phase shift of Titan's contrast relative to the seasonal forcing. Describing the contrast in terms of a relative normal albedo, we calculate the time dependence of Titan's disk-integrated brightness and find a significant seasonal contribution. Finally, we use the spectral and phase angle dependence to place constraints on the mechanism which produces the contrast.

Titan's brightness contrast at the time of Voyager 1

We selected nearly simultaneous Voyager narrow angle frames in violet (0.37–0.45 μm), blue (0.43–0.53 μm), and green (0.52–0.60 μm) filters (A. Collins, P. Kupferman and E. Danielson, personal communication), pictures 1305S1-003, 1289S1-003, and 1287S1-003 respectively. These were taken at 2.38×10^6 km from Titan, at a phase angle of 29.6°, with a resolution of 44 km per line pair at the subspacecraft point. The Voyager blue and green intensities roughly correspond to Lockwood's measurements with the b filter (0.45–0.48 μm) and y filter (0.54–0.56 μm) respectively[3]. Relative intensities were obtained from raw digital counts by subtracting an empirically determined dark noise correction accurate to ~ ±0.5 count.

To define limb darkening characteristics each image was scanned along the 150° E meridian, using a 5×5 pixel area average stepped at ~3° intervals (Fig. 1). These scans pass between the subspacecraft point at 134 °E and the subsolar point at 162° E, providing nearly maximal variation in scattering geometry and intensity. Figure 1 shows the dependence of intensity I (in digital counts) on μ (cosine of the observer zenith angle) and μ_0 (cosine of the solar zenith angle) as $I\mu$ versus $\mu\mu_0$ on logarithmic scales so that Minnaert scattering behaviour, for which

$$I\mu = I_0(\mu\mu_0)^k \qquad (1)$$

is a straight line with slope equal to the limb darkening exponent k, and intercept (at $\mu\mu_0 = 1$) equal to the maximum intensity I_0. Each scan follows two nearly linear traces with similar slopes, moving along one line in the northern hemisphere and a second line, with an intercept ~25% greater, in the southern hemisphere. A transition occurs near the equator. Because roughly the same limb darkening slope applies over most of the latitude range for each colour, we can interpret brightness differences as differences in the intercept I_0 (proportional to the normal albedo).

We define a latitude dependent relative normal albedo (Fig. 2) as the ratio of the observed intensities to the intensities predicted from a Minnaert function fit to the 150° E scan data between 31° and 52° S for each colour. The fitted slopes are 0.773, 0.887, and 0.942 for violet, blue, and green respectively. Four major cloud zones are apparent in Fig. 2: a north mid-latitude zone between about 10° and 55° N, a north polar zone between 55° and 90° N, a south mid-latitude zone between about 20° and 60° S, and an equatorial transition zone between about 20° S and 10° N. The large variations poleward of 60° S and 80° N cannot be interpreted as true variations in normal albedo as they occur where the determination of the viewing geometry is very sensitive to navigation errors (Fig. 1) and where the Minnaert function does not accurately describe Titan's limb darkening. The normal albedo of the south mid-latitude zone is ~25% greater than that of the north mid-latitude zone, while the albedo of the north polar zone is ~7% smaller. The greatest normal albedo contrast between the south and north mid-latitude zones occurs in the blue, and the smallest in the green.

The large-scale spatial structure of the normal albedo, as well as subtler cloud features are displayed as normalized images (Fig. 3) in which the intensity is proportional to relative normal albedo. The images clearly show the major cloud zones, and indicate a slightly brighter band at 45° N as identified by Smith *et al.*[7]. The features in the normalized images are primarily zonal (axially symmetric). The north polar zone is dark in all three colours, but darkest in violet. The bright band just south of this zone is most pronounced in the green and nearly invisible in the

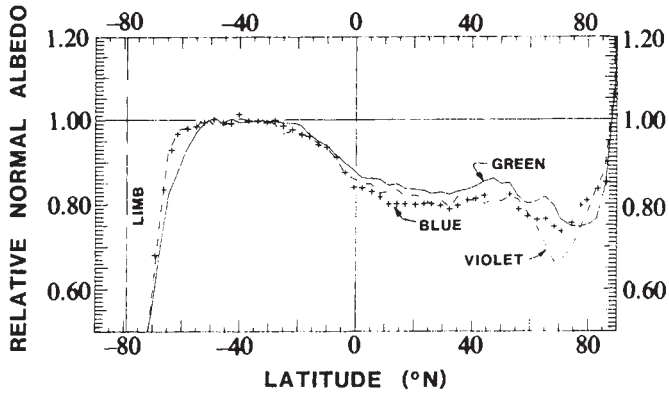


Fig. 2 Relative normal albedo versus latitude along the 150° E meridian. For each colour the relative normal albedo is obtained by dividing observed intensities by predictions based on Minnaert fits to the intensities between 31 and 52° S. The increase of normal albedo north of 70° N is not real: a decrease similar to that near 70° S is found when μ and μ_0 in the polar zone are corrected for the larger height of the clouds there. The decrease near the limb in the south is similar to that found near the limb at other latitudes and is probably a failure of the Minnaert fit to describe the true limb darkening at large viewing angles.

© 1981 Nature Publishing Group

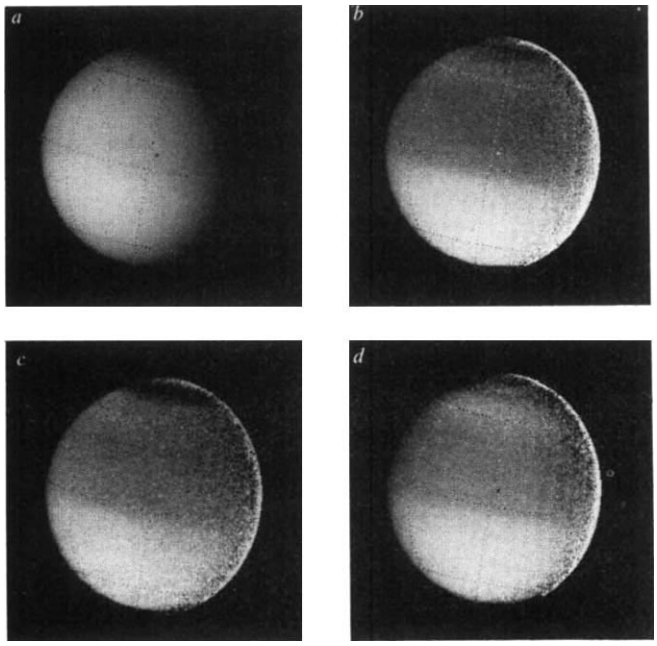


Fig. 3 Voyager 1 narrow angle images of Titan at 29.6° phase angle: unnormalized blue image (*a*) and normalized images in blue (*b*), violet (*c*), and green (*d*). In all the normalized images intensities are linear in relative normal albedo (see text) but stretched to enhance the visibility of the relatively small contrast range. A similar linear stretch has been applied to the unnormalized image (*a*). Its different appearance is primarily due to the presence of limb darkening. The latitude–longitude grid overlay has a 45° spacing. The equator intersects the 135° E meridian near the centre of the image (+, Titan's centre). The bright terminator in the normalized images probably results from the haze layer above the main cloud deck[7], which could cause the observed intensities to fall off more slowly than the predicted intensities, yielding an apparent albedo increase.

violet. Minor, but definite, longitudinal variations also occur. The north polar zone boundary, for example, is slightly inclined relative to latitude lines. The green image also shows an inclined bright feature within the polar zone which may be useful as a tracer of atmospheric motions. Those small departures from axial symmetry have an insignificant effect on Titan's disk-integrated brightness and are ignored.

Phase angle dependence

We assume that the normal albedo contrast is the same at 0° phase angle (corresponding to Earth-based observing geometry) as it is at the 29.6° phase angle applicable to the Voyager 1 data shown in Figs 1–3. To test this assumption we analysed a Voyager 1 blue filter image taken at a phase angle of 9.3° (picture 1379S1–014), using the same procedures as in the analysis of the 29.6° phase angle images. No discernible change in mid-latitude contrast was found between these phase angles. The small size of the 9.3° phase angle image (30 picture elements across Titan's diameter) precludes a definitive statement about polar region changes.

Time dependence of Titan's contrast

The possible time dependence of Titan's brightness contrast can be constrained by comparing the Voyager 1 and Pioneer 11 observations (1.17 yr apart). We used a blue filtered image made by the Pioneer 11 imaging photopolarimeter (IPP) at a phase angle of 27° and at a surface resolution of ~170 km at the subspacecraft point (M. G. Tomasko, personal communication). The spectral range of the IPP blue filter (0.39–0.50 μm) covers both the blue and violet bands of the Voyager camera system. Because we see little spectral dependence of contrast in this spectral range, and as the albedo of Titan is higher in the blue than in the violet, we expect no significant spectral error in comparing the IPP image with the Voyager blue image.

Using the same procedures as above, a central meridian scan of the IPP image was used to define the relative normal albedo as a function of latitude. The large quantization noise in the IPP data was reduced by averaging over two latitude belts in which the normal albedo seemed to be constant: 15°–35° S and 30°–60° N. We found a south/north normal albedo ratio of 1.26 ± 0.03 for the Pioneer measurements, compared with 1.25 ± 0.01 for Voyager. This result is consistent with either a fixed contrast or a slowly varying contrast near an extremum. We reject the former possibility because it implies brightness variations grossly inconsistent with the Earth-based photometry. The Pioneer 11/Voyager 1 comparison also suggests that the normal albedo boundary (the point at which the albedo is midway between the mid-latitude values) has moved from about 5° N to 5° S, apparently opposite to the subsolar point. But this tentative result is not included in our seasonal model calculations.

If the contrast slowly varies with a seasonal period (29.5 yr), then the direct observations by Voyager 1 and Pioneer 11 place a strong constraint on the phase shift between the solar forcing due to spin axis tipping and the contrast response. If both the forcing and the contrast response vary sinusoidally, the implied phase shift is −84° ± 20°, where −90° corresponds to a time delay of one season, and the implied maximum south/north ratio of normal albedos is $1.26^{+0.05}_{-0.01}$. The relatively large uncertainty in phase is due to the small time interval between Voyager and Pioneer and the large quantization noise in the Pioneer results. Present results are consistent with maximum contrast near the time of the Voyager 1 encounter and allow for a large enough phase lag (near 90°) to place the contrast-induced brightness variation in phase with the Earth-based photometry.

If the normal albedo response to the solar irradiance level can be described by a single time constant, then the phase shift is determined by $\tan(\phi) = -2\pi\tau/T$, where ϕ is the phase shift, T is the orbital period of Saturn, and τ is the time constant. As $\phi \leq -90°$ is not physically possible, the time constants consistent with the phase shifts allowed by the Pioneer 11/Voyager 1 comparison fall in the range $9.6\ \text{yr} \leq \tau \leq \infty$, with the central phase shift of 84° corresponding to a time constant of 44.7 yr. This range of time constants is not consistent with a direct interaction between solar irradiance and aerosol properties at the unit optical depth level (at a pressure of ~1 mbar), because the thermal radiative time constant at this level and the time to reach equilibrium particle properties[8] are both <1 yr. Large time constants can be obtained, however, if deeper levels of the atmosphere are involved in the contrast producing mechanism. For the entire atmosphere down to the 1.6-bar level a time constant of 138 yr (ref. 7) $\phi = -88°$ is possible. A time constant of 9.6 yr only requires involvement down to the 100 mbar level.

Predicted geometric albedo time dependence

We now evaluate the time dependence of Titan's disk-integrated brightness expected for two models consistent with the Pioneer 11/Voyager 1 comparison: fixed contrast and seasonal contrast with a large phase shift (−90° for this example). Because the Earth-based observations of Titan's disk-integrated brightness[1,2] have been adjusted to fixed heliocentric distance and zero phase angle (180° scattering angle), they are proportional to Titan's geometric albedo (the ratio of disk-integrated brightness to that of a flat perfect Lambertian reflector of the same diameter). We calculate the relative geometric albedo $p(t)$ by integrating over the illuminated hemisphere:

$$p(t) = (1/\pi)\iint a(\theta, t)[\mu_0(\theta, \theta_s(t), \phi)]^{2k} \sin(\theta)\, d\theta\, d\phi \quad (2)$$

where $a(\theta, t)$ is the relative normal albedo at co-latitude θ and time t, θ_s is the co-latitude of the subsolar (and subEarth) point, and ϕ is the longitude relative to the subsolar longitude. The time dependence of the subsolar point latitude on Titan is obtained by assuming Titan's spin axis is parallel to Saturn's. We approximate the latitude dependence of the normal albedo using a 16-zone step function with boundaries spaced ~10° in latitude. Average relative albedo values for each zone are calculated for each colour from the results shown in Fig. 2. Because of the navigational and Minnaert fit uncertainties we

© 1981 Nature Publishing Group

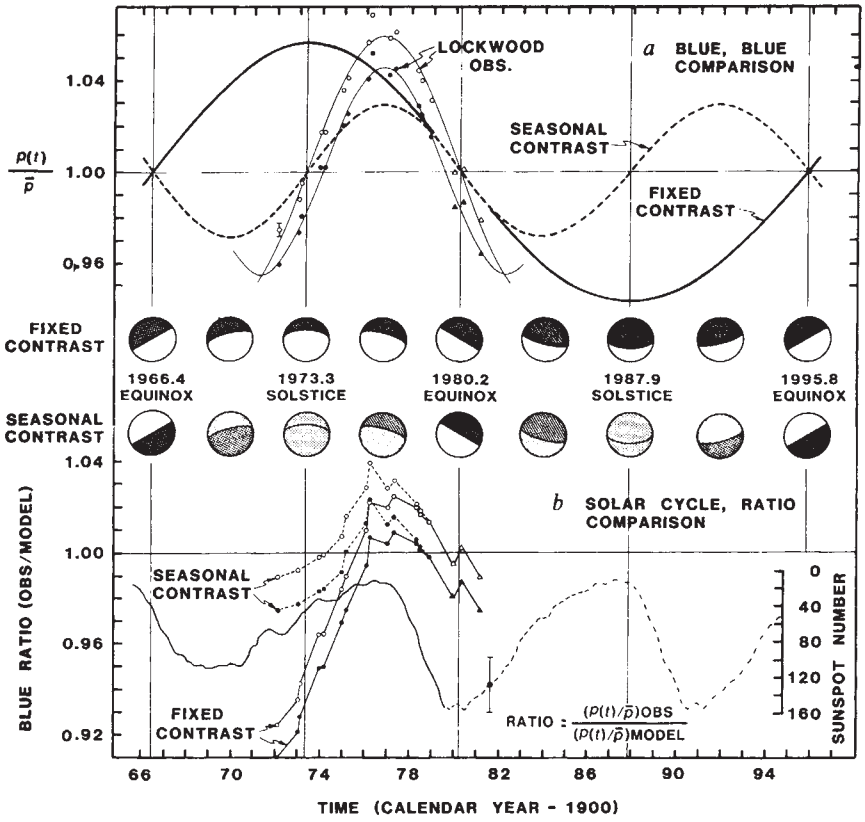


Fig. 4 *a*, Relative time variation of Titan's blue geometric albedo according to fixed contrast (heavy solid line) and seasonal contrast (heavy dashed line) models. Relative time variations derived from Lockwood's Earth-based photometry are shown for two different estimates of the mean brightness, each based on fitting an offset sine function of fixed period to the observations. One case uses the 11-yr solar cycle period; the second uses the current 13.8-yr period of the predicted seasonal brightness variation (the ellipticity of Saturn's orbit causes a deviation from the mean 14.7-yr period). Both fits (light lines) are good: r.m.s. deviations from the fitted curves are only about 1.5 times the estimated measurement error. Observations: ○△, relative to the mean assuming a 13.8-yr period; ●▲, relative to the mean assuming an 11-yr period. Triangles indicate observations which are still preliminary. The shaded drawings illustrate the time-dependent appearance of Titan at zero phase angle according to the two contrast models. The drawings exaggerate and simplify the contrast for clarity. *b*, Comparison of the solar cycle of smoothed sunspot number[11] with the ratio of observed to predicted relative variations in Titan's blue geometric albedo. Sunspot numbers are plotted along an inverted scale. The dashed portion of the sunspot curve indicates predicted behaviour based on previous cycles. The error bar indicates short term prediction uncertainty. Beyond 1983 the uncertainty is comparable with the predicted value.

assume the same albedo south of 60° S as between 52.5° S and 60° S, and ignore variations north of 85° N in finding a mean between 70° and 90° N. This normal albedo distribution is valid for 1980.9, the time of the Voyager 1 encounter. The fixed contrast model assumes the same distribution at all times. We describe possible seasonal variations in normal albedo in a simple model assuming (1) the normal albedo distribution for 1980.2 (the time of the last equinox) was the same as observed by Voyager for 1980.9, (2) the distribution for 1966.4 (time of the previous equinox) was as for 1980.2 except for a north–south reversal, and (3) at each latitude the normal albedo varied sinusoidally between the 1966.4 and 1980.2 extremes.

For fixed albedo contrast we find peak-to-peak modulation amplitudes of ~10.5% for violet and blue geometric albedos, and 8.2% for the green. For seasonally-varying contrast the corresponding values are 5.6% and 4.2% respectively. Figure 4*a* illustrates the time dependence of the geometric albedo based on the normal albedo models for the blue intensities. For fixed albedo contrast, $p(t)$ varies in phase with the subsolar motion, reaching extreme values at the solstices. For the seasonal contrast model, $p(t)$ attains its mean value at every equinox, when both hemispheres are equally illuminated and viewed, and at every solstice, when the interhemispheric normal albedo contrast is absent. In this case the extreme values occur midway between solstice and equinox. The shaded drawings (Fig. 4*a*) provide a simplified illustration of Titan's time dependent appearance according to these two contrast models.

Comparison with Earth-based photometry of Titan

The Lockwood measurements of Titan's b magnitude are plotted in Fig. 4*a* as relative brightnesses (brightness/time-averaged brightness). The most recent observations should be considered preliminary (G. W. Lockwood, personal communication).

The brightness variation calculated assuming a fixed normal albedo distribution is completely at odds with the Earth-based photometry. The predicted brightness is at a maximum in 1973 and beginning to decline, while the measurements are near the estimated mean value and rapidly increasing. This implies that the mean normal albedo of Titan's clouds and the relative lbedo distribution with latitude cannot both be time-independent. If the relative spatial distribution of normal albedo were fixed, then the disk-averaged mean value would have to vary dramatically to compensate as in Fig. 4*b*, which shows the ratio of Lockwood's blue observations to the predictions. As the ratio (for fixed contrast) changes most dramatically before 1976 when solar indicators (such as sunspot number, Fig. 4*b*) are changing relatively slowly, the possibility of fixed contrast and a large compensation by the solar UV mechanism seems highly unlikely.

If the relative latitudinal variation of normal albedo has the seasonal model time dependence, the predicted geometric albedo variation has roughly the same modulation period and phase as the Earth-based measurements. The assumed seasonal phase lag of 90° provides nearly optimal agreement with the Earth-based b brightness variation (Fig. 4*a*), although the optimal phase lag for fitting both b and y brightness curves is 82°±7°, where the uncertainty is based on the difference between best fit values for the two colours. The predicted peak-to-peak modulation amplitude based on the green contrast measurements is ~50–70% of the observed y brightness modulation, while the predicted blue variation is about 50–60% of the observed b brightness modulation, where the amplitude uncertainties are due to the uncertainties in Titan's mean brightness. For agreement with the Earth-based photometry the mean normal albedo can be modulated by only ~5% in blue and only 2.5% in the green (or y filter). This implied variation in disk-averaged mean normal albedo (Fig 4*b*) is roughly in phase with that expected from the solar cycle modulations of UV output (Fig. 4*b*), although the magnitude of the solar cycle effect in the present case would be only about half the amount suggested before Titan's brightness contrast had been discovered.

Implications of Neptune's brightness variation

The seasonal contribution to Titan's brightness variation could be significantly different from the above estimate if the boundary between high and low albedo also exhibited seasonal variation: it would be greater if the boundary moved opposite to the subsolar point motion (as suggested by Pioneer 11 measurements) and smaller if it moved in the same direction.

©1981Nature Publishing Group

Thus dominance by either solar cycle or seasonal contributions remains possible. Earth-based photometry of Neptune is helpful in evaluating this possibility.

As Neptune's orbital period of 164.8 yr seems to preclude any significant seasonal changes during the 1972–80 period of the Earth-based photometry, we might expect its brightness curve to be dominated by the solar UV mechanism. The asymmetric form of Neptune's brightness variation supports this expectation: after relatively smooth increases between 1972 and 1976.5, at an average rate of 0.54% yr^{-1} in the blue and 0.68% yr^{-1} in the yellow, the brightness curves for both colours dropped by more than 2% in the yellow and almost 3% in the blue within the next year, with relatively little additional change from 1978 to 1980 (ref. 2 and G. Lockwood, personal communication). Thus Neptune's brightness curves are relatively well correlated with the asymmetric variation of the indicators of solar activity (Fig. 4*b*).

If the shape of Neptune's brightness curve represents the relative time dependence of the solar cycle contribution to Titan's brightness, then the solar cycle mechanism cannot be dominating Titan's variation. The very symmetric form of Titan's brightness curve is instead consistent with dominance by the seasonal mechanism. However, between 1972 and 1976 Titan's b magnitude increased ~70% faster than the y magnitude, while between 1976 and 1980 they both changed at about the same rate. We interpret this spectral difference as evidence for a non-seasonal component. From the amplitude of Neptune's brightness variation we infer that the solar cycle component in Titan's brightness curve would contribute ~30% of the total variation. Although one might expect a larger effect because Titan's haze layer is optically thick, while Neptune's is thought to be optically thin, a contribution much greater than 30% should produce a noticeable asymmetry in Titan's brightness curve. Thus a simple seasonal component probably accounts for more than half of Titan's secular variation, with the remainder being provided by the solar cycle mechanism.

Constraints on the contrast mechanism

As most photons reflected to space by Titan interact with the atmospheric aerosols, the hemispheric brightness asymmetry is almost certainly due to variations in their properties. The phase angle dependence of the brightness asymmetry gives an indication as to which property of the aerosols has a dominant role. Theoretical simulations with the aerosol parameters of Rages and Pollack[9] indicate that the contrast declines considerably and even disappears as the phase angle decreases from 30° to 10° for a model involving variations in mean particle size, whereas little contrast alteration is expected over the same range of phase angles for models involving variations in aerosol absorption coefficient or optical depth. Thus the latter models seem to be in better agreement with the observed lack of significant contrast variation with phase angle.

The spectral dependence of the contrast also suggests that it is a reflection of differences in the sub-micrometre aerosol properties rather than differences in albedo of an underlying methane cloud layer. If the contrast was due to an albedo difference deep in the atmosphere it would be attenuated by aerosols in the upper atmosphere. However, to explain the spectral variation of Titan's geometric albedo the visible absorption coefficient of these particles must increase significantly with decreasing wavelength[9]. As their inferred optical depth is higher at shorter wavelengths their attenuation of the north–south contrast of an underlying cloud layer should result in significantly more contrast at long rather than short wavelengths. As this is opposite to the observed spectral variations, the contrast is more likely to be present in the properties of the sub-micrometre particles in the upper atmosphere.

The large phase shift of the contrast relative to the seasonal forcing functions shows that the contrast-producing mechanism cannot be a direct interaction between the solar irradiance and the atmosphere at the unit optical depth level. A significant depth of Titan's atmosphere must be involved to produce the time delayed response. As a relatively small north–south thermal contrast is expected at levels corresponding to large radiative time constants, a direct thermal effect on the composition or optical depth of the aerosol particles seems unlikely. The observed vertical stability of Titan's temperature profile[10] also argues against an interhemispheric difference in small-scale vertical convection as a modifying influence. This suggests that large-scale circulation asymmetries may be responsible. Possibly differences in meridional circulations, driven by an interhemispheric seasonal thermal contrast, are affecting the aerosol properties by changing the mix of trace gases available at the formation level, or by modifying their rates of removal from the formation region.

Summary and conclusions

Titan's normal albedo distribution is primarily axially symmetric with a substantial north–south contrast. As the contrast is essentially independent of phase angle it is likely to result from differences in the composition or optical depth of Titan's layer of sub-micrometre aerosols rather than differences in size distribution. The spectral variation of the contrast suggests that it is not caused by differences in the albedo of an underlying methane cloud. The contrast seems to vary with time in response to seasonal variations in solar irradiance and lags the seasonal forcing by a large phase angle, indicating that it originates probably well below the unit optical depth level. It seems to be produced indirectly, perhaps by circulation currents driven by thermal contrast at lower levels which affect the rate of removal or composition of the aerosol particles at the unit optical depth level. The contrast has a significant effect on Titan's disk-integrated brightness, and is probably responsible for most of the observed long term variation, with solar cycle effects accounting for the remainder.

There are still many uncertainties regarding the nature and origin of the contrast. A longer record of observations, both of the contrast and the disk-integrated brightness would yield a significant improvement in understanding. The nature of the contrast variation should be better defined by Voyager 2 observations in August 1981, and by the Space Telescope in the mid-1980s. If the simple seasonal model is correct we would expect the contrast to decrease to zero by the end of 1987.

Continued Earth-based photometry of both Titan and Neptune is also essential to resolve the remaining uncertainties. Between 1972 and 1978 the solar cycle component and the seasonal component have been approximately in phase, and neither has yet experienced a full cycle of variation, making their separate contributions difficult to distinguish from the brightness curves alone. The two components will become readily separated with another 5–10 years of observations as they grow increasingly out of phase, moving in opposite directions between 1981 and 1984.

We thank B. A. Smith and the Voyager imaging team for support, S. A. Collins and J. A. Anderson for help in obtaining the necessary Voyager data, G. W. Lockwood for recent brightness measurements for Titan and Neptune, M. G. Tomasko for normalized Pioneer 11 observations of Titan, and T. Wendricks for drafting the figures. This research was partially supported by JPL contract 953615.

Received 15 June; accepted 9 July 1981.

1. Lockwood, G. W. *Icarus* **32**, 413–430 (1977).
2. Lockwood, G. W. & Thompson, D. T. *Nature* **280**, 43–48 (1979).
3. Suess, S. T. & Lockwood, G. W. *Solar Phys.* **68**, 393–409 (1980).
4. Vidal-Madjar, A. in *The Solar Output and its Variation* (ed. White, O. R.) 213–236 (Colorado Associated University Press, Boulder, 1977).
5. Cook, J. W., Brueckner, G. E. & Van Hoosier, M. E. *J. geophys. Res.* **85**, 2257–2266 (1980).
6. Pollack, J. B., Rages, K. & Toon, O. B. *Geophys. Res. Lett.* **7**, 829–832 (1980).
7. Smith, B. A. *et al. Science*, **212**, 163–191 (1981).
8. Toon, O. B., Turco, R. P. & Pollack, J. B. *Icarus* **43**, 260–284 (1980).
9. Rages, K. & Pollack, J. B. *Icarus* **41**, 119–130 (1980).
10. Hanel, R. *et al. Science* **212**, 191–200 (1981).
11. *Solar Geophysical Data, prompt reports*, No. 425 (US Dept of Commerce, 1980).

© 1981 Nature Publishing Group

Density waves in Saturn's rings

Jeffrey N. Cuzzi*, Jack J. Lissauer† & Frank H. Shu†

*Ames Research Center, NASA, Moffett Field, California 94035, USA
†Department of Astronomy, University of California, Berkeley 94720, California, USA

Certain radial brightness variations in the outer Cassini division of Saturn's rings may be spiral density waves driven by Saturn's large moon Iapetus, in which case a value of ~16 g cm^{-2} for the surface density is calculated in the region where the waves are seen. The kinematic viscosity in the same region is ~170 cm^2 s^{-1} and the vertical scale height of the ring is estimated to be a maximum of ~40 m.

VOYAGER imaging observations of Saturn's rings[1] have revealed a wide spectrum of radial structure, indicating a variety of dynamical processes in the ring disk. Here we focus on one example of radial structure and explore its possible origin. A sequence of smoothly undulating brightness fluctuations is observed in the outer Cassini division within a band located between ~120,700 and 121,900 km from the centre of Saturn. The observed behaviour of these variations agrees with the expected behaviour of spiral density waves driven by a resonance between the apsidal motion of local ring particles and the mean motion of Saturn's outermost large satellite, Iapetus.

Basic theory

The existence of spiral density waves driven at Lindblad resonances with Saturn's major moons and propagating radially in Saturn's ring disk was suggested by Goldreich and Tremaine[2] as a means of transferring angular momentum over large radial distances within the disk and thereby clearing wide gaps. Similar density waves have been proposed to underlie the spiral structure of disk galaxies[3]. The earlier work was extended by Goldreich and Tremaine[4,5] to the density waves that can be driven in thin disks by external and internal satellites.

Consider an axisymmetric disk of ring particles orbiting Saturn where centrifugal equilibrium at a radius r requires circular motion at angular speed $\Omega(r)$. If the particles possess small random motions in the plane of the disk (small eccentricities), Lindblad's theory of orbits[6] allows us to think of the particles as gyrating at epicyclic frequency $\kappa(r)$ about the guiding centre of the reference circular orbit of the same specific angular momentum $r^2\Omega(r)$. Then

$$\kappa^2 = \frac{1}{r^3}\frac{\mathrm{d}}{\mathrm{d}r}(r^4\Omega^2) \tag{1}$$

with κ^2 being equal to Ω^2 if the latter satisfies Kepler's third law: $\Omega^2 \propto r^{-3}$. For Saturn's rings, κ^2 is slightly less than Ω^2 because of the oblateness of Saturn. The small departure of κ^2 from Ω^2 is crucial to the nature of the resonant driving of the observable spiral density waves.

Suppose a moon of mass M orbits Saturn *outside* the rings at a mean angular speed Ω_M. Its perturbing force on the ring particles can be Fourier decomposed in time t and azimuthal angle θ (ref. 2). The mth Fourier component, having dependence proportional to $\exp[im(\Omega_M t - \theta)]$ with m being a positive integer, will resonate with ring particle motions at a radius r_L where

$$\Omega(r_L) - \kappa(r_L)/m = \Omega_M. \tag{2}$$

Following the nomenclature used in galactic dynamics, we call the value r_L which satisfies equation (2) the mth inner Lindblad resonance of the moon M.

Ring particles slightly inside and outside r_L are induced to travel in eccentric orbits with longitude of periapse differing by π/m, leading to overlapping of their orbits. For $m > 1$, collisions would then depopulate the overlapping orbits and clear natural gaps about the stronger resonances with fractional widths $\Delta r/r_L \sim (M/M_S)^{1/2}$, where M_S is the mass of Saturn. Applied to Mimas's 2:1 resonance (the $m = 2$ inner Lindblad resonance), the above formula fails by more than two orders of magnitude to explain the observed width of the Cassini division.

To account for this discrepancy and to explain why Mimas's 2:1 resonance occurs near the inner edge of Cassini's division, Goldreich and Tremaine[2] invoked the self-gravity of the ring material. Consider free spiral density waves, of arbitrary wave frequency ω and azimuthal symmetry m, in a flat disk of particles which collectively behave as a fluid. Independent of their source, these waves satisfy the dispersion relation[7]:

$$(\omega - m\Omega)^2 = \kappa^2 - 2\pi G\sigma|k| + c^2k^2 \tag{3}$$

where σ is the surface mass density, c is the one-dimensional velocity dispersion, and k is the radial wavenumber in a WKBJ analysis. Equation (3) states that the allowed square of the Doppler-shifted wave frequency equals the square of the natural frequency (the epicyclic frequency) apart from modifications required by self-gravity and acoustic effects.

From equation (3), it is easy to show that stability to axisymmetric ($m = 0$) disturbances of all wavenumbers requires the fluid analogue of Toomre's criterion[8]:

$$Q \equiv \frac{\kappa c}{\pi G\sigma} \geqslant 1 \tag{4}$$

When $Q \geqslant 1$, equation (3) predicts that the disk can support two types of free density waves with

$$|k| = \frac{\pi G\sigma}{c^2}[1 \pm (1 - Q^2D/\kappa^2)^{1/2}] \tag{5}$$

where the upper sign corresponds to short waves and the lower to long waves, and where the departure of

$$D \equiv \kappa^2 - (\omega - m\Omega)^2 \tag{6}$$

from zero measures how far we are from exact Lindblad resonance. The value of r where

$$\omega - m\Omega(r) = -\kappa(r) \tag{7}$$

is called the inner Lindblad resonance of the wave of radian frequency ω and azimuthal mode number m. For a wave driven by an external moon of pattern speed $\omega/m = \Omega_M$, the radial position corresponding to equation (7) is $r = r_L$ (equation (2)).

The waves of interest here are the long waves. Near the inner Lindblad resonance of the wave, $Q^2D/\kappa^2 \ll 1$, and the dispersion relation for long waves becomes independent of c (see, for example, equation (3) for small $|k|$):

$$|k| = \frac{D}{2\pi G\sigma} \tag{8}$$

Equation (8) predicts that these waves formally become infinitely long at the waves' inner Lindblad resonance where $D = 0$. Because the characteristic wavelength far from resonance, $4\pi^2G\sigma/\kappa^2 (\simeq 4\pi H/Q$ where $H \equiv c/\Omega$ is the scale height of the disk), amounts only to ~10^2 m in Saturn's rings, it is the reduction of the wavenumber $|k|$ by the factor D/κ^2 that allows the waves at inner Lindblad resonance to be coupled to the slow spatial variation of the forcing by the external moon[2,3]. The long waves—in particular the $m = 1$ long waves—have the added observational advantage that they can be resolved by Voyager 1.

0028-0836/81/340703—05$01.00

© 1981 Macmillan Journals Ltd

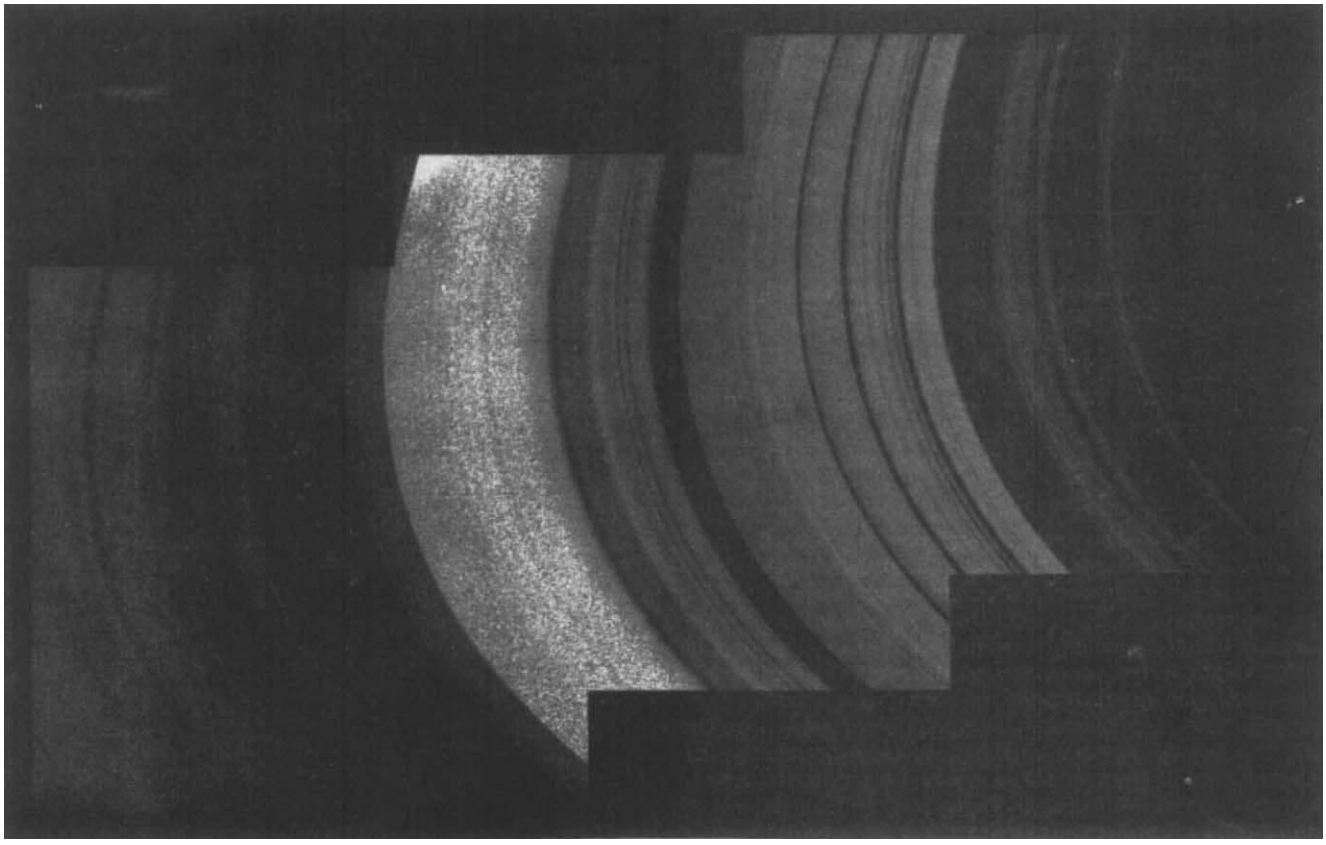


Fig. 1 A mosaic of the Cassini division of Saturn's rings, obtained from Voyager 1 observations with resolution of ~10 km per line pair (see Table 1). The region is seen from the unlit face. The prominent bright band in left centre contains a sequence of features of alternating brightness which are plotted in Fig. 2.

To understand why the $m = 1$ inner Lindblad resonances (where the rate of apse precession of the ring particles is equal to the mean motion of the moon) are of special interest, we determine the radial range over which the waves remain significantly longer than the characteristic wavelength of 10^2 m. Because $D = 0$ at $r = r_L$ for driven waves, a Taylor series expansion of D about $r = r_L$ for driven waves yields

$$D = \mathscr{D}x \tag{9}$$

where

$$\mathscr{D} \equiv \left(r\frac{\mathrm{d}D}{\mathrm{d}r}\right)_{r_L} = \left(2\kappa r\frac{\mathrm{d}}{\mathrm{d}r}(\kappa - m\Omega)\right)_{r_L} \tag{10}$$

and $x \equiv (r - r_L)/r_L$ is the fractional distance from Lindblad resonance. Equation (8) now implies that the radial spacing, or wavelength $\lambda = 2\pi/|k|$, between successive wave crests should be inversely proportional to the nondimensional distance x from the resonance:

$$\lambda = \frac{4\pi^2 G\sigma}{\mathscr{D}x} \tag{11}$$

with the proportionality constant being larger for smaller values of the parameter $\mathscr{D}$. But in Saturn's rings, κ nearly equals Ω, so equation (10) now implies that the most readily resolved density waves correspond to $m = 1$. Of the larger exterior moons, the only ones whose $m = 1$ inner Lindblad resonance lies within the ring system are Iapetus, Hyperion, and Titan.

In the $m = 1$ case, to calculate $\mathscr{D}$ (but not to locate the resonance), it is sufficiently accurate to take

$$\kappa - m\Omega \simeq -J\left(\frac{GM_S}{r^3}\right)^{1/2}\left(\frac{R_S}{r}\right)^2 \tag{12}$$

where $J = 0.0244$ and $R_S = 60{,}330$ km are, respectively, Saturn's dynamical oblateness and equatorial radius. This yields for equation (10):

$$\mathscr{D} \simeq 7J\Omega_L^2(R_S/r_L)^2 \tag{13}$$

Let us consider the amplitude distribution of the propagating waves whose source is $m = 1$ resonant driving at $r = r_L$ by an external moon. Free long waves that propagate conserving the angular momentum luminosity F given them by a source can be shown[2,9] to increase their fractional amplitude linearly with dimensionless distance x from the inner Lindblad resonance, $|\Delta\sigma/\sigma| = x/x_{NL}$, where x_{NL} is the characteristic scale for waves to become nonlinear:

$$x_{NL} = \frac{2\pi^2}{\mathscr{D}}\left(\frac{r\sigma}{F}\right)^{1/2}(G\sigma)^{3/2} \tag{15}$$

For weak forcing, as occurs with Iapetus on Hyperion, $x_{NL} \gtrsim 1$, and nonlinear effects (for example, shocks[10]) are not important for realistic values of x. In the presence of particle collisions, however, small-amplitude density waves will suffer a viscous damping factor[2]

$$\exp\left[-\int_0^x k_I r_L\,\mathrm{d}x\right] \qquad \text{where } k_I = \frac{7}{3}\left[\frac{\kappa\mathscr{D}^2}{(2\pi G\sigma)^3}\right]\nu x^2$$

if the bulk viscosity is much less than the kinematic shear viscosity ν. Taking into account both amplitude variation factors, we obtain the wave amplitude profile near the position of the inner Lindblad resonance:

$$\frac{\Delta\sigma}{\sigma} = \left(\frac{x}{x_{NL}}\right)\exp\left[-\left(\frac{x}{x_D}\right)^3\right] \tag{16}$$

where

$$x_D \equiv \left(\frac{9}{7}\right)^{1/3}\frac{2\pi G\sigma}{(\kappa\nu\mathscr{D}^2 r)^{1/3}} \tag{17}$$

is the characteristic dimensionless damping length.

If spiral density waves are observed in Saturn's rings, a measurement of their wavelength yields, through equation (11), a determination of the surface mass density σ; a measurement of the fractional amplitude variation then yields, through equations (16) and (17), a determination of the kinematic viscosity ν. Under certain assumptions, we can relate the kinematic viscosity ν to ring optical depth τ and velocity dispersion c. The vital parameters σ and ν are difficult to obtain empirically by other methods.

© 1981 Nature Publishing Group

Table 1 Observational parameters of Cassini bright band

Image FDS counts used	34934.25, 34934.29
Resolution	10 km per line pair
Ring tilt angle to line of sight	74°
Normal optical depth	0.14 ± 0.02
Fit to wavelength: $\lambda(x) = A_1/(x - x_1)$	$A_1 = 0.45 \pm 0.07$ km
	$x_1 = -0.0001 \pm 0.0012$
Fit to amplitude $\frac{\Delta\sigma}{\sigma} = A_2 x\, e^{-(x/x_2)^3}$	$A_2 = 1.69 \pm 0.2$
	$x_2 = 0.0115 \pm 0.001$
Theoretical amplitude $A_2 = (x_{NL})^{-1}$	1.3 (using $M/M_S = 4 \times 10^{-6}$)
Derived σ	16 ± 3 g cm^{-2}
Derived ν	170 ± 80 cm^2 s^{-1}

Observations and derived ring properties

During the Voyager 1 encounter with Saturn[1], several observations of the Cassini division region were obtained with spatial resolution of ~10 km per line pair (see Table 1). As viewed from the unlit face, an optically thin region of ~4,800 km width is observed between the classical A and B rings; we refer to this entire region as the Cassini division (see also ref. (11)). In the outer part of the Cassini division, at a distance from Saturn's centre of ~120,700–121,900 km (±700 km) is a noticeably brighter band filled with regular, low contrast, radial brightness variations having spacings that decrease outward (see Fig. 1). Independent of the high resolution observations shown in Fig. 1, the band was observed against the bright limb of Saturn and the average normal optical depth of the band was determined as discussed by Smith *et al.*[1] (see Table 1). The band appears brighter in diffuse transmission than nearby regions in the inner Cassini division because the band's optical depth is a factor of ~1.4 larger than the plateau structures in the inner Cassini division.

To obtain useful results from the very noisy data evident in Fig. 1, an azimuthal averaging technique was employed, in which data were averaged along loci of constant distance from the centre of Saturn. This method employs numerical fitting of ellipses to observed ring features at the boundaries of the region. To first order, geometrical distortion in the camera and due to the finite separation between the spacecraft and the planet is thereby removed. Typical azimuthally-averaged results, which incorporate most of the data contained in the frame, are shown in Fig. 2. The vertical brightness scale has been calibrated to reflectivity $DN = I/F$, in units of 10^{-4}, where I is the observed intensity and πF is the incident solar flux. The noise level of the result may be estimated from inspection of the darkest area, which is essentially empty. In particular, the mildly undulating variations located between 120,700 and 121,900 km from the centre of Saturn are real.

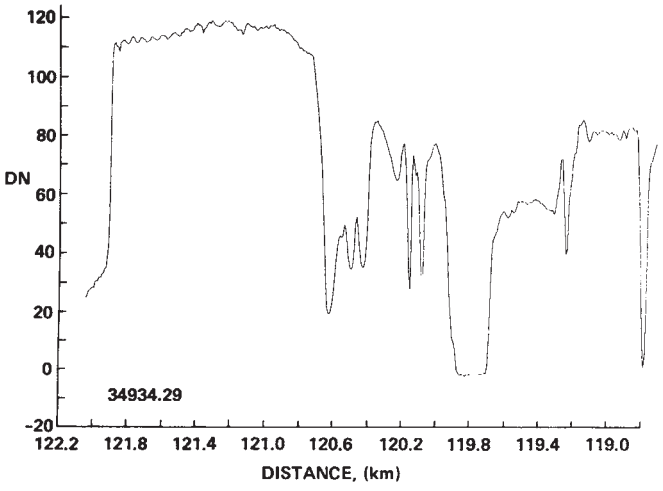


Fig. 2 Photometric scan across part of the data shown in Fig. 1. Averaging is done along lines of constant distance from Saturn, and the tracing includes most of the information contained in Fig. 1. The noise level is shown in the dark (empty) gap near 119,800 km. Radial distances shown are only accurate in an absolute sense to ±700 km or so, but relative distances are much more precisely known.

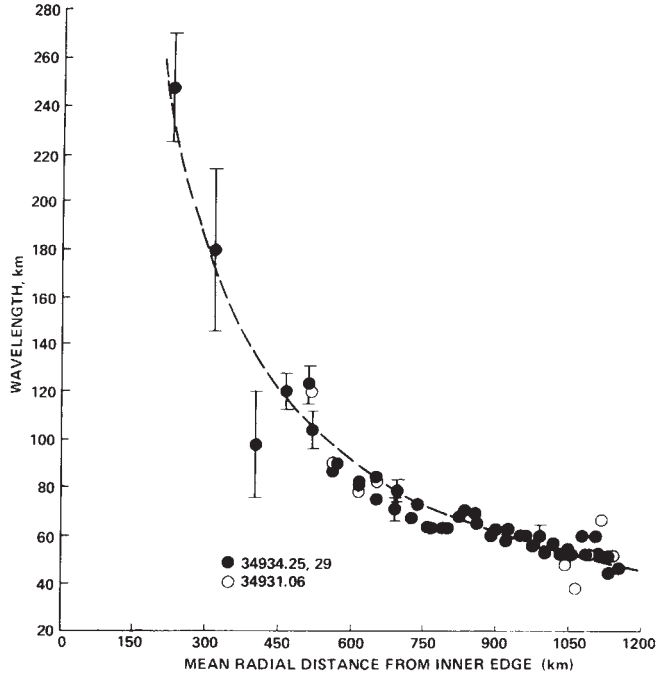


Fig. 3 Wavelength, or peak–peak/trough–trough separation, of the wave features of Figs 1 and 2 as a function of mean distance from the inner half-power point of the bright band. The best-fit curve of the form of equation (11) is shown by the dashed curve. The inverse variation of wavelength with distance is characteristic of long density waves near Lindblad resonance.

The horizontal distance scale in Fig. 2 is obtained from final spacecraft trajectory data and contains residual uncertainties of ±700 km for this observation. This uncertainty is large compared with the uncertainty <100 km to which the theoretical radial location of a given gravitational resonance may be calculated. However, relative distances are quite accurately represented in Fig. 2, and therefore, resonances can be located relative to one another once one of them has been identified.

We have calculated the predicted locations and strengths of the strongest resonances in this region of the rings following the approach of Goldreich and Tremaine[5]. The strongest resonance is the $m = 2$ inner Lindblad resonance with Mimas. An $m = 1$ Lindblad resonance with Iapetus lies at 2.008 R_S, but is several orders of magnitude weaker in torque. If we associate the Mimas 2 : 1 resonance with the inner edge of the Cassini division (which might entail a slight error[11]), relative separations locate the Iapetus resonance at 100 ± 150 km outward of the inner half-power point of the brightest band in Fig. 2. Because the undulations in brightness start near this location a more detailed investigation of the hypothesis that these undulations represent a radial slice through a tightly-wrapped spiral density-wave pattern whose source is the Iapetus resonance is necessary.

We have analysed the azimuthally averaged data obtained from two independent frames for the wavelength and amplitude behaviour of the wave pattern. For the low optical depths characterizing the region, we equate fractional brightness variations with fractional surface density variations. The fractional separations $x = (r - r_0)/r_0$ of the centroids of the peaks and troughs from an arbitrary radial location r_0 (the inner half-power point of the bright band) were measured. The peak-to-peak as well as trough-to-trough wavelengths were then plotted as functions of the distance from this arbitrary reference point. The results are shown in Fig. 3 with a best fit to the data of the form of equation (11):

$$\lambda = \frac{A_1}{x - x_1} \tag{18}$$

where A_1 and x_1 were varied to minimize residuals. The resulting zero point offset $x_1 = -0.0001 \pm 0.0012$ implies that the source of the observed wave pattern is indistinguishable from the inner half-power point of the bright band. This is slightly inward of the theoretical location deduced on the basis of identifying the inner edge of the Cassini division with Mimas's

© 1981 Nature Publishing Group

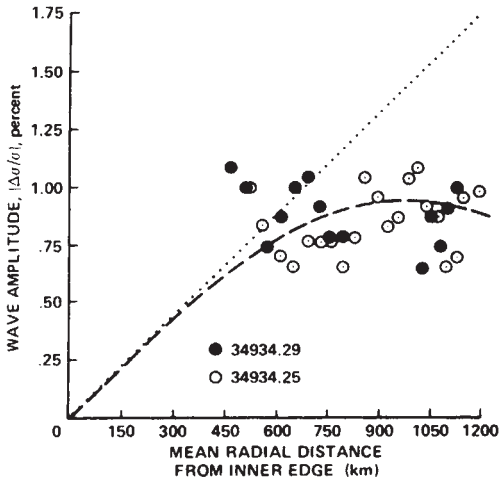


Fig. 4 Wave amplitude data (1/2 of peak–trough contrast) are plotted for the wave features of Figs 1 and 2 as a function of distance from the inner edge of the bright band. These data have been fit to a function of the form of equation (16), and the best-fit curve is shown by the dashed line. In the absence of damping, the amplitude would continue to increase linearly throughout the region instead of leveling off and then declining.

$m = 2$ inner Lindblad resonance, but the displacement is within the allowed uncertainties.

Comparing equations (11) and (18), we see that determination of A_1 allows us to calculate the local mass surface density σ. Our best-fit result yields $\sigma = 16 \pm 3$ g cm^{-2}. This is the most secure determination of a local ring property from our analysis and constitutes, to our knowledge, the only direct estimate for the ring mass density. The normal optical depth τ and surface mass density are related by

$$\tau = K\sigma \tag{19}$$

where K is the opacity, or mean cross-sectional area per unit mass of ring particles of density ρ:

$$K(r) = \frac{\int n(R, r)\pi R^2 \, dR}{\int n(R, r)\rho(4\pi R^3/3) \, dR} \tag{20}$$

with $n(R, r)$ representing the local distribution of particle radii R. Independent measurements of τ and σ yield a constraint on the ratio of integrals, K. Our analysis gives $K = \tau/\sigma \approx (9 \pm 2) \times 10^{-3}$ cm^2 g^{-1} for the region of interest.

For comparison, for identical particles of radius R_0: $n(R, r) \propto \delta(R - R_0)$, $K = 3/4\rho R_0$ whereas, for a power-law distribution: $n(R, r) \propto R^{-3}$ from R_1 to R_2, as has been deduced from ground-based studies[12], $K = [3/(4\rho(R_2 - R_1))] \ln (R_2/R_1)$. Voyager radio science results[13] for the region of our analysis indicate an effective particle radius $R_{\rm eff} \sim 8$–10 m. If this corresponded to identical particles of radius $R_0 \sim R_{\rm eff}$ and $\rho \sim 1$ g cm^{-3}, it would imply $K \sim 9 \times 10^{-4}$ cm^2 g^{-1}. In contrast, the power-law distribution, with $R_1 \sim 1$ cm (ref. 12) and $R_2 \sim R_{\rm eff}$ (ref. 13) would predict $K \sim 6 \times 10^{-3}$ cm^2 g^{-1}, in much better agreement with our determination. Our analysis therefore supports a power-law distribution with an exponent of ~ 3, although the precise value of the exponent depends on our choice for R_1 and R_2. In particular, a steeper spectrum would allow a larger value of R_2, possibly including many thousands of kilometre-sized bodies within the ring system[14].

We now investigate whether the observed amplitude profile of the wave is consistent with the behaviour predicted by equation (16). There are two separate parameters, $x_{\rm NL}$ and $x_{\rm D}$, to be considered. First, the best-fit value for $x_{\rm NL}$ using equation (16) or the data of Fig. 2 may be compared with the expected value obtained from equation (15) using a theoretical calculation for the strength F of the Iapetus apsidal resonance and the observational determination of σ given above. Second, the best-fit for $x_{\rm D}$ using equation (16) may be used to obtain an estimate for the kinematic viscosity ν from equation (17).

The data for fractional wave amplitudes from our two pictures are shown in Fig. 4 with the minimum-residual fit for a functional form given by equation (16). Although the data exhibit considerable scatter the waveform for density waves damped by viscosity does provide an acceptable fit to the observations. Moreover, the derived value of $x_{\rm NL}$ is within a factor of 1.3 of the theoretically expected value for wave excitation and propagation in a region of constant σ (see Table 1). This slight discrepancy may be accounted for by uncertainties in Iapetus' mass or by noting that the resonant source probably lies in a region of larger surface mass density than the region of wave propagation. Within these uncertainties, Iapetus apsidal resonance appears to contain sufficient torque to account for the observed wave amplitude.

We proceed to estimate the value of ν from our best fit for the damping length $x_{\rm D}$. Using equation (16) and our value of σ, we obtain $\nu \sim 170$ cm^2 s^{-1} (see Table 1). Although this value is less precise than our value for σ, the presence of appreciable viscosity is unquestionable given the clear departure seen in the data of Fig. 4 from an undamped linear increase of fractional wave amplitude with radial distance from the resonance.

As the waves suffer damping, the negative angular momentum they carry is absorbed by the ring particles, causing the latter to drift inward towards the inner Lindblad resonance. But this process itself does not seem to explain why the region containing the waves is a region of lower optical depth than its surroundings, as the total angular momentum transfer induced by Iapetus over the entire age of the Solar System fails by two orders of magnitude to produce the required mass removal. We believe the question of the low optical depth of the outer part of the Cassini division to be related to that of the origin and maintenance of the low optical depth of the entire Cassini division[11].

Under certain assumptions, the derived value of ν may be used to obtain the particle velocity dispersion c and thus the local vertical scale height $H = c/\Omega$. For instance, for particles of a single size, Goldreich and Tremaine[15] have derived a form for ν which can be approximated by

$$\nu = \frac{c^2}{2\Omega}\left(\frac{\tau}{1+\tau^2}\right) \tag{21}$$

This expression neglects gravitational focusing and scattering[16], and is valid only if the distribution of ring particles contains few bodies with radii much in excess of 10 m. (The free fall speed onto a 10-m particle is comparable with the value of c we derive below using equation (21).) As equation (21) ignores contributions to the viscosity other than physical collisions, our determination of c based on ν yields only an upper limit. If we adopt equation (21) and set $\nu \sim 170$ cm^2 s^{-1} and $\tau \sim 0.14$, we obtain $c \leqslant 0.6$ cm s^{-1}, corresponding to a ring scale height $H = c/\Omega \leqslant 40$ m. From the values for σ and c, we obtain $Q \leqslant 30$ for the stability index, equation (4). The upper limit $Q = 30$ is still small enough to justify our use of the approximation $Q^2 D/\kappa^2 \approx Q^2 \mathscr{D}x/\kappa^2 \ll 1$ to derive equation (8) for the region in question.

Discussion

The above results have several general implications. If ν has everywhere the same value, 170 cm s^{-1}, as it does in the outer part of Cassini's division, the unconstrained spreading time $\Delta r^2/\nu$ of Saturn's rings as a whole would be comparable with the age of the Solar System. Smaller features would, of course, be much more difficult to maintain for the age of the Solar System. Indeed, unless the coefficient in the ν–τ relation decreases rapidly with increasing optical depth (see equation (21)), ν is likely to be several times larger in the B and A rings than in the low optical depth region studied here. The persistence of even these gross features over the age of the Solar System would be called into question without another confinement mechanism.

In contrast to constant ν or c, a value for Q near its upper limit 30 everywhere would imply that the collective effects of the self-gravity of the rings are generally unimportant except very near a resonance. In the case of the $m = 1$ waves, the waves are extremely close in frequency space to exact Lindblad resonance. In general, however, density waves become evanescent beyond

© 1981 Nature Publishing Group

the distance from their source where equation (5) produces complex wavenumbers. This location, known in galactic dynamics as a 'Q barrier'[17], occurs in close spatial proximity to the source for all $m \neq 1$ in Saturn's rings unless Q is close to unity. For example, for density waves originating at the Mimas $m = 2$ inner Lindblad resonance to propagate across even a continuously filled region as wide as Cassini's division, Q must be <3. If Cassini's division was cleared by such density waves[15], Q and therefore c have to be so small that the primary contribution to the viscosity comes from gravitational scatterings by massive ring particles. In any case, our measured value of ν does require some means of maintaining the present Cassini division at a low surface density against radial diffusion by the B and A rings in a time $\Delta r^2/\nu \leqslant 4 \times 10^7$ yr.

The Titan apsidal resonance at a theoretical radial location of 1.288 R_S is also orders of magnitude stronger in torque than the Iapetus apsidal resonance, but unlike the Mimas $m = 2$ resonance, the Titan $m = 1$ Lindblad resonance is free of difficulties with Q barriers, and the waves it excites should be resolvable. Yet we see no evidence, in high resolution studies, of density waves near 1.288 R_S, only a gap of a few hundred km width exists[18].

We believe that the seemingly paradoxical presence of observed density waves associated with the weak Iapetus resonance and their apparent absence from the much stronger Titan resonance has a simple explanation in the very strength of the latter resonance. It and the other strong resonances are so strong that their driven density waves become nonlinear in a short distance[2], $\Delta r = x_{NL} r_L$, and shock[10], depositing most of their negative angular momentum in close proximity to the resonance, and thereby quite likely opening up a gap. The distance to nonlinearity for both the Mimas 2 : 1 resonance and the Titan apsidal resonance is 10^2 km, as compared with 10^5 km for the weak Iapetus apsidal resonance. Goldreich and Tremaine[2] pointed out that the behaviour of the wave becomes highly uncertain after the material within $x_{NL} r_L$ of the resonance has been removed, and they also expressed concern in the case of the Mimas 2 : 1 resonance as to why the entire A ring had not been drawn into the Cassini division.

The presence of an empty region to the outside of a resonance diminishes the direct coupling of the satellite's resonant forcing to the disk on the outer side of the gap by an exponential factor if the waves must tunnel through a smoothly varying Q-barrier (edges not sharper than an inverse wavenumber). This factor in the amplitude may be roughly estimated from equation (53) of ref. 17, in the limit of small σ, as

$$\exp\left[-\frac{2}{3}\left(\frac{\mathscr{D}\Delta r}{\kappa^2 r}\right)^{1/2}\frac{\Delta r}{H}\right]$$

where Δr is the observed width of the gap. The magnitude of the exponent is >10 for the observed gap near Titan's apsidal resonance and even larger for the strong nonapsidal resonances in the rings. Thus, direct inward transport of material by strong resonances with outer moons may be self limiting. If density waves play a major part in transporting matter in Saturn's rings, it must be possible to excite waves at nonresonant positions via interactions with preexisting small-scale structure (Goldreich, personal communication).

Thus strong evidence exists in Saturn's rings for satellite-driven density waves, one in particular being driven by Iapetus at its apsidal resonance. Based on this identification, we obtain values for ring properties such as $\sigma = 16\ \mathrm{g\,cm^{-2}}$; $K = \tau/\sigma = 9 \times 10^{-3}\ \mathrm{cm^2\,g^{-1}}$; $\nu = 170\ \mathrm{cm^2\,s^{-1}}$; while $c \leqslant 0.6\ \mathrm{cm\,s^{-1}}$, $H \leqslant 40$ m, and $Q \leqslant 30$. Firm *lower* bounds for c and H can be obtained by requiring axisymmetric gravitational stability when $Q \geqslant 1$, then our value of σ gives $c \geqslant 0.02\ \mathrm{cm\,s^{-1}}$, $H \geqslant 1.5$ m. Narrowing the ranges for c, H, and Q requires a careful evaluation of the relative contributions to ν by gravitational scatterings and physical collisions. This will depend critically on whether the power-law distribution of ring particles extends to bodies >10 m.

We thank our colleagues on the Voyager imaging team for their support and help, especially B. A. Smith, A. Collins, L. Soderblom, A. Bunker, and G. Dimit. We also thank P. Goldreich and S. Lubow for helpful conversations, and K. Bilski, M. Leroy, and K. Fischer for assistance with data reduction. This work was supported in part by NASA grant NC A2-OR 050-102.

Received 4 June; accepted 28 July 1981.

1. Smith, B. A. *et al. Science* **212**, 163 (1981).
2. Goldreich, P. & Tremaine, S. *Icarus* **34**, 240 (1978).
3. Lin, C. C. & Shu, F. H. *Astrophys. J.* **140**, 646 (1964).
4. Goldreich, P. & Tremaine, S. *Astrophys. J.* **233**, 857 (1979).
5. Goldreich, P. & Tremaine, S. *Astrophys. J.* **241**, 425 (1980).
6. Chadrasekhar, S. *Principles of Stellar Dynamics* (University of Chicago Press, 1942).
7. Lin, C. C. & Shu, F. H. in *Astrophysics and General Relativity* Vol. 2 (eds Chretian, M., Deser, S. & Goldstein, J.) 236–329 (Gordon and Breach, New York, 1968).
8. Toomre, A. *Astrophys. J.* **139**, 1217 (1964).
9. Shu, F. H. *Astrophys. J.* **160**, 99 (1970).
10. Shu F. H., Milione, V. & Roberts, W. W. *Astrophys. J.* **183**, 819 (1973).
11. Lissauer, J. J., Shu, F. H. & Cuzzi, J. N. *Nature* **292**, 707–711 (1981).
12. Cuzzi, J. N. & Pollack, J. B. *Icarus* **33**, 233 (1978).
13. Tyler, G. L. *et al. Science* **212**, 201 (1981).
14. Henon, M. *Nature* (in the press).
15. Goldreich, P. & Tremaine, S. *Icarus* **34**, 227 (1978).
16. Cuzzi, J. N., Durisen, R., Burns, J. A. & Hamill, P. *Icarus* **38**, 54 (1979).
17. Lin, C. C. & Lau, Y. Y. *Stud. appl. Math.* **60**, 97 (1979).
18. Collins, S. A. *et al. Nature* **288**, 439 (1980).

Moonlets in Saturn's rings?

Jack J. Lissauer*, Frank H. Shu* & Jeffrey N. Cuzzi†

*University of California, Berkeley, California 94720, USA
†NASA/Ames Research Center, Moffett Field, California 94035, USA

The brightness structure within Cassini's division in Saturn's rings is explained in terms of perturbations produced by moonlets embedded within an optically thin disk of smaller ring particles. The moonlets exert gravitational torques on neighbouring ring particles and create gaps; diffusion acts to fill the gaps. A new explanation is offered for the inner edge of the Cassini division being located at the 2 : 1 resonance with Mimas.

VOYAGER 1 observations have revealed that Saturn's rings consist of a surprisingly large number of ringlets and narrow gaps[1]. Some of the features in the rings appear to be associated with resonant forcing by external moons[2]. We have argued that some observed undulating variations of small amplitude in the outer Cassini division can be interpreted as propagating spiral density waves whose source is resonant driving by Iapetus[3]. However, there also exist ring structures which cannot be easily explained in terms of external gravitational perturbations.

Ring particles typically have sizes of 0.01–10 m (ref. 4), and Voyager 1 discovered several small new moons with diameters of 30–200 km just outside the main ring system[1]. Thus, we suspected the existence of bodies of intermediate size, and we investigated the idea that some of the prominent structures in Saturn's ring system might be caused by the gravitational influence of such relatively massive bodies orbiting within the ring system (see also ref. 5).

To have enough cohesive strength to survive inside the Roche limit of Saturn, the perturbing bodies must have diameters $\leqslant 100$ km (ref. 6). To exert enough gravitational torque to clear gaps larger than their own sizes, the bodies must have diameters $\geqslant 2$ km (see equation (14)). We refer to bodies embedded in the

0028–0836/81/340707—05$01.00

© 1981 Macmillan Journals Ltd

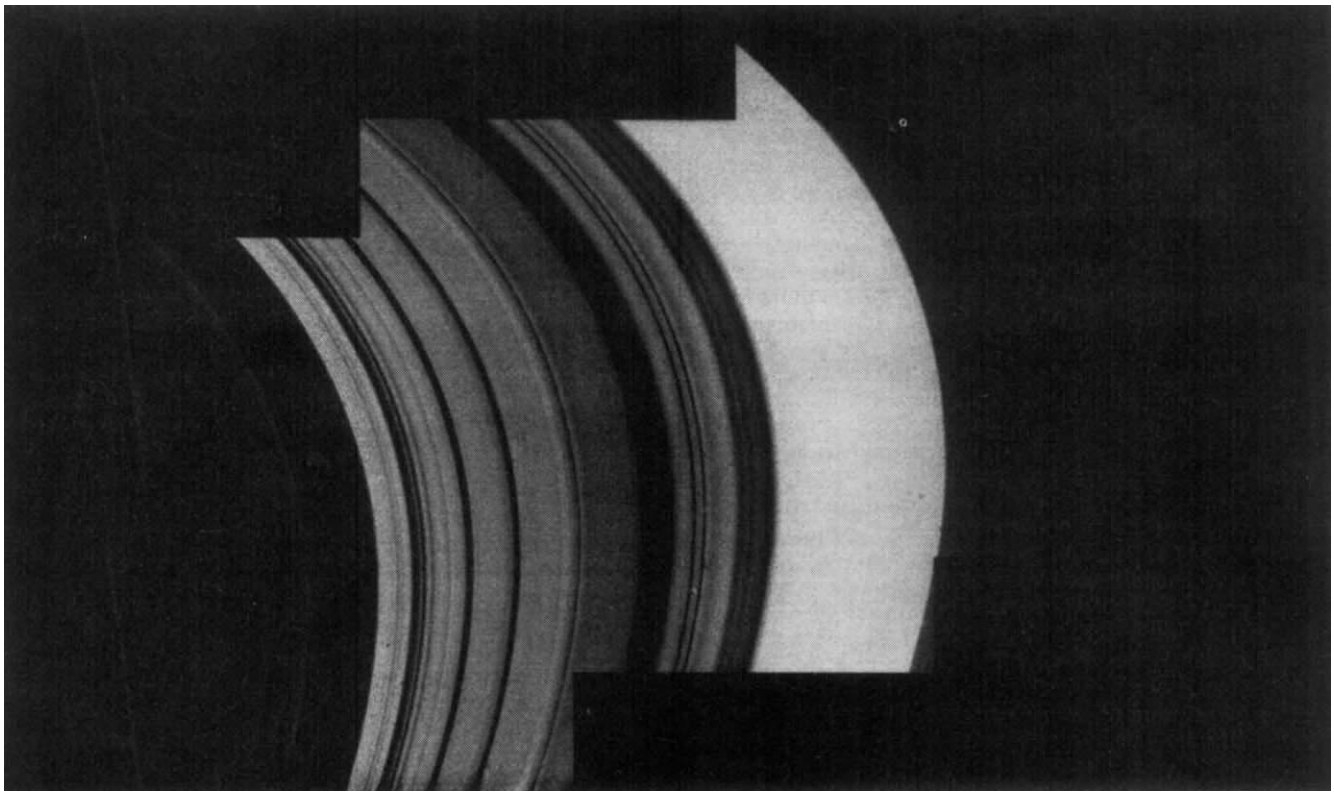


Fig. 1 Mosaic of the Cassini division in diffuse transmission. The B ring at the left and the A ring at the right appear dark because the relation, diffuse brightness proportional to optical depth, breaks down when the multiple scatterings occur. The inner Cassini division contains a series of plateaux separated by three dark narrow gaps and bounded by two dark wide rifts. The outer Cassini division (between the outer rift and the A ring) contains a faint wavelike pattern that is analysed in ref. 3.

rings which are massive enough to clear gaps larger than their own sizes as moonlets. We aim here to obtain mass estimates for unseen moonlets that may be responsible for the clearest gaps and to analyse whether such bodies could be directly detected by future spacecraft observations. We focus on the series of plateaus in the inner Cassini division which are separated by narrow gaps and set off from the B ring and the outer Cassini division by two wide gaps or 'rifts' (see Fig. 1).

Our analysis of this region follows that of ring dynamics given by Goldreich and Tremaine[7,8] and by Lin and Papaloizou[9]. After the discovery of the rings of Uranus[10], Goldreich and Tremaine proposed an ingenious mechanism for creating and maintaining narrow planetary rings[11]. Their mechanism invoked small unseen satellites between the rings that keep the ring particles from spreading under the influence of collisions. The discovery of two satellites[1], 1980 S26 and 1980 S27, which seem to 'shepherd' Saturn's narrow F ring supports the hypothesis.

A moonlet embedded within the ring system will exert a systematic torque on ring material located at the moonlet's outer and inner Lindblad resonances, where the ratio of the mean revolution rates of the ring Ω and of the moonlet Ω_M satisfy $\Omega/\Omega_M = m/(m \pm 1)$, with m as a positive integer. The direction of this torque always acts to move ring material away from the moonlet. Collisions between ring particles broaden the resonances, and near the orbit of the moonlet, the Lindblad resonances of large m pile up and overlap. The torques can then be considered as spaced continuously over the ring material. Each moonlet thus sweeps ring material on both sides away from itself. If the moonlets are large or closely spaced, particles between the two moonlets may be pushed into quite a narrow ringlet, as is the case with the uranian rings and Saturn's F ring. Counteracting this effect, viscous diffusion tends to maintain the ringlet at a finite width. If the relative role of diffusion is increased—by an increase in the amount of material between adjacent moonlets or in the random velocities of the ring particles, by a decrease in the masses of the moonlets, or by an increase in their separation—the ringlets can spread and become plateaux wider than the gaps between them. We suggest this occurs for the prominent plateau structures in Fig. 1.

Basic equations

In a ring system where the normal optical depth τ is of the order unity, collisions between ring particles occur on time and length scales of the ring orbital period and thickness. For structures of longer duration and larger dimension, we may treat the ring material as a viscous fluid, confined to a thin sheet. Ignoring the effects of gravitational focusing and scattering[12], and assuming a single particle size, Goldreich and Tremaine[13,14] proposed a formula for the kinematic viscosity that we approximate with

$$\nu = \frac{c^2}{2\Omega}\left(\frac{\tau}{1+\tau^2}\right) \tag{1}$$

where c is the one-dimensional velocity dispersion of ring particles. The visual optical depth τ is related to the surface mass density σ and mean cross-sectional area per unit mass of ring particles, 'opacity' K, by

$$\tau = K\sigma \tag{2}$$

In contrast, if the size distribution of embedded bodies extends from centimetres to kilometres (or larger) on a single-power law of sufficient flatness, a more important source of effective viscosity than physical collisions arises from the tendency of the larger bodies to exert torque on their neighbours in such a way as to repel the latter[5]. The same mechanism operates on all size scales, but a natural separation occurs at the size R_2 that separates bodies large enough to clear gaps around themselves (moonlets) from bodies too small to do so (ring particles). We treat the latter collection as a fluid with an effective kinematic viscosity ν given by dimensional arguments as

$$\nu = \alpha R_2 \frac{G\sigma}{\Omega} \tag{3}$$

where α is a dimensionless constant of the order of unity whose exact value depends on the exponent of the power-law that characterizes the ring-particle size-distribution.

© 1981 Nature Publishing Group

Although equations (1) and (3) have different origins, they both give ν proportional to σ for small τ. This is the only important feature for our modelling of the inner Cassini divisions as the coefficient of proportionality between ν and τ or σ can be determined empirically. We therefore adopt equation (1) with the understanding that the quantity c should be regarded, not as the true velocity dispersion of ring particles, but merely as a parameter in the ν–τ relation. Further, we assume c to have a spatially constant value $\sim 0.5\ \mathrm{cm\ s^{-1}}$, given approximately by our determination of ν through the damping of spiral density waves in the outer Cassini division[3]. An equivalent procedure would have been to take $\alpha R_2 = 0.2$ km in equation (3); virtually identical results would then have been obtained in Fig. 2.

The tangential stress exerted by material outside radius r (distance from Saturn) on the material inside radius r, integrated over the thickness of the rings, is $\sigma\nu r\, \partial\Omega/\partial r$. This stress has a lever arm r and acts (in an axisymmetric system) over a circle of circumference $2\pi r$. The net torque exerted on a narrow annulus is the difference between these effects at $r + \mathrm{d}r$ and r; therefore, the viscous torque T_ν exerted per mass unit is given by[15]

$$2\pi r\sigma T_\nu = \frac{\partial}{\partial r}\left(2\pi r r\sigma\nu r\frac{\partial\Omega}{\partial r}\right) \tag{4}$$

In the present problem, we can ignore the self-gravity of the rings and the oblateness of Saturn; we approximate $\Omega(r)$ by Kepler's third law; and equation (4) becomes

$$T_\nu = -\frac{3}{2r\sigma}\frac{\partial}{\partial r}(\sigma\nu r^2\Omega) \tag{5}$$

Moonlets clear gaps around themselves by exerting torques on

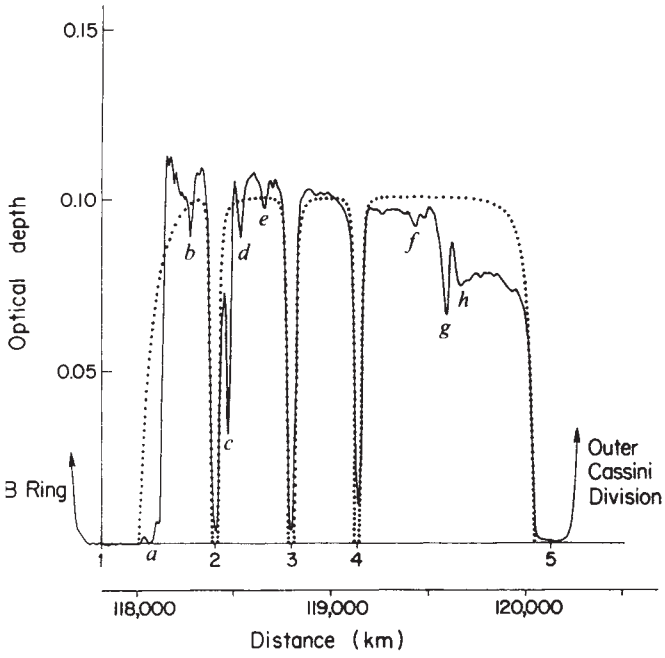


Fig. 2 Comparison of the theoretical optical depth profile with the measured diffuse transmission in the inner Cassini division. The solid curve gives the observed elliptical ring scans as a function of distance from Saturn; the dotted curve gives a theoretical fit to the data. The theoretical model adopts constant velocity dispersion, two stationary rifts moonlets, 1 and 5, and three gap moonlets, 2, 3 and 4, which have evolved to quasi-equilibrium positions. The moonlets have the properties listed in Table 1. Additional smaller moonlets may exist at positions a–h, but they were not modelled. In particular, moonlet a is required to confine the small bump (which is an optically thick narrow ring) between moonlets 1 and a, and to sharpen the inner edge of the first plateau. In these medium-resolution scans the optical depth does not reach zero at the positions 2–4, but this is an artefact as high-resolution scans show that clear gaps exist at these positions. Equilibrium solutions were also calculated for c increasing with τ. Resultant profiles have sharper gap edges, as expected from the slower diffusion rates in the lowest optical depth regions. If c decreases slowly with increasing τ, rounder edges will result. Rapid decrease of c with τ can result in diffusive instabilities[20].

Table 1 Properties of the moonlets

Moonlet number j	M_j/c (g cm^{-1} s)	D_j (km)*
1	2.5×10^{19}	29
2	7.0×10^{17}	9
3	7.0×10^{17}	9
4	5.2×10^{17}	8
5	6.9×10^{18}	19

* The calculation for D_j from M_j/c assumes $\rho = 1$ g cm^{-3} and $c = 0.5$ cm s^{-1}.

ring particles in various kinds of resonances. If the eccentricity of a moonlet of mass M_j and radial position $r_j(t)$ is small, the torque $T_j(r, t)$ it exerts on a unit mass of fluid at radius r can be obtained from equation (18) of ref. 8, specialized to a keplerian velocity field:

$$T_j = 0.40\ \mathrm{sgn}\,(r - r_j)\left[\frac{GM_j}{\Omega(r - r_j)^2}\right]^2 \tag{6}$$

Once moonlets have emptied gaps larger than their own sizes, we may ignore accretion by and erosion of moonlets. Mass conservation of ring particles leads to the equation of continuity

$$\frac{\partial\sigma}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}(r\sigma u) = 0 \tag{7}$$

where u is the azimuthally averaged radial fluid velocity of the ring particles. Given the separate conservations of mass of ring particles and moonlets, we write the equations governing the transfer of angular momentum between the ring fluid and the J moonlets as

$$\frac{\mathrm{D}}{\mathrm{D}t}(r^2\Omega) = T_\nu + \sum_{j=1}^{J} T_j \tag{8}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}[r_j^2\,\Omega(r_j)] = -\frac{1}{M_j}\int_0^\infty T_j(r, t)2\pi r\sigma(r, t)\,\mathrm{d}r + T_j^{\mathrm{ext}} \tag{9}$$

where $\mathrm{D}/\mathrm{D}t \equiv \partial/\partial t + u\,\partial/\partial r$ is the substantial derivative and T_j^{ext} is the specific torque exerted on the jth moonlet because of possible resonant locking to a massive external moon[16]. In equation (8), we have assumed an absence of direct perturbations on the ring material by external moons; and in equation (9), we have ignored the possibility of resonant moonlet–moonlet interactions. The latter interactions are small for the moonlet masses and separations implied by the ring structure in Cassini's division.

In the keplerian approximation, $\partial(r^2\Omega)/\partial t = 0$ and $\partial(r^2\Omega)/\partial r = r\Omega/2$; thus, equation (8) becomes

$$\tfrac{1}{2}ur\Omega = T_\nu + \sum_{j=1}^{J} T_j \tag{10}$$

Given c, Ω, K, M_j and T_j^{ext}, equations (1), (2), (5), (6), (7), (9) and (10) form a closed set to solve for ν, τ, T_ν, T_j, σ, r_j and u. We now show how quasi-equilibrium conditions simplify the solution to these equations.

Approximate equilibrium behaviour

From equations (10), (5) and (6), we readily estimate the fluid velocity u with which the jth moonlet will clear a ring gap Δr where the viscous torque has a magnitude comparable with that of the moonlet's gravitational torque:

$$ur\Omega \sim \frac{\nu r\Omega}{\Delta r} \sim \left[\frac{GM_j}{\Omega(\Delta r)^2}\right]^2 \tag{11}$$

so the characteristic time scale Δt for the gap clearing process is

$$\Delta t \sim \frac{\Delta r}{u} \sim \frac{(\Delta r)^2}{\nu} \sim \frac{(GM_j)^{4/3}}{\Omega^2 r^{2/3}\nu^{5/3}} \tag{12}$$

which, with typical ring properties[3], is of the order of several years for a moonlet of 2 km diameter, and scales as the moonlet diameter to the fourth power. Therefore, moonlets randomly placed among ring fluid will clear gaps around themselves in a time short in comparison with the age of the Solar System. The

© 1981 Nature Publishing Group

ring particles must quickly move to accommodate a near balance between the torques on the right-hand side of equation (10). Approximating the spatial variation of c, Ω, r and K to be small in comparison with the resulting distribution of τ, we may set the right-hand side of equation (10) to zero and obtain

$$\frac{\tau}{1+\tau^2}+\arctan(\tau)=-\frac{0.18}{r\Omega^2}\sum_{j=1}^{J}\frac{(GM_j/c)^2}{|r-r_j|^3}+C_0 \qquad (13)$$

where C_0 is a dimensionless integration constant. For ring material far from any moonlets, $C_0=\tau_0/(1+\tau_0^2)+\arctan(\tau_0)$, and equation (13) predicts an optical depth profile which takes the form of a plateau ($\tau=\tau_0$), with a steep decline as r approaches the position of the nearest moonlet on either side of the plateau. The edge of the gap, $\tau=0$, is reached at a value of r where the right-hand side of equation (13) is zero. Equation (13) is not valid beyond this point.

For given local disk properties, there is a minimum size of a body which can lead to a clear gap. This minimum moonlet size is found by requiring that the equilibrium gap width exceed the size of the body embedded within the gap. With a spherical moonlet of internal density ρ, from equation (13) the criterion for the diameter D of a moonlet is

$$D \gtrsim 1.4(C_0 r)^{1/3}\left(\frac{\Omega c}{G\rho}\right)^{2/3} \qquad (14)$$

where C_0 is a monotonic function of the plateau optical depth τ_0. For small τ_0, as in Cassini's division, C_0 is of the order of $2\tau_0$. With values which apply to Cassini's division today[3], equation (14) implies that D must exceed roughly 2 km.

If an unconstrained moonlet ($T_j^{\text{ext}}=0$) were to find itself between two plateaus of different τ_0 (and therefore σ), the right-hand side of equation (9) will have a sign that forces this moonlet to move away from the high-density plateau towards the low-density plateau. The high-density plateau then spreads by viscous diffusion while the low-density plateau is compressed by the approaching moonlet. This reduces the density difference $\Delta\sigma$ between the two plateaus. When $\Delta\sigma$ has been reduced sufficiently and the evolving gap has developed enough width, the motion of the moonlet will be quite slow in comparison with the time required to produce the quasi-equilibrium ring profile given by equation (13) for fixed moonlet positions.

This suggests a simple procedure for computing the slow phase of the evolution of Saturn's ring system (most of the age of the Solar System). We calculate the moonlets' radial velocities by substituting the fluid distribution between adjacent moonlets implied by equation (13) into equation (9). New quasi-equilibrium fluid distributions can then be calculated from equation (13) after the moonlets are moved, keeping the amount of ring material constant between each pair of moonlets (Fig. 2).

Unless the boundaries of (part of) the disk are constrained by external effects (for example, moonlets locked to heavy exterior moons), no true equilibrium is possible for the ring system. (A monolayer with $c=0$ and $\sigma\neq 0$ would not spread, but it would be unstable to collective gravitational disturbances.) If the inner and outer boundaries correspond to moonlets, there will be no rings to either side, and unbalanced torques in equation (9) will force these moonlets to move away from the middle. Similarly, if the inner and outer boundaries correspond to ring fluid, there will be no moonlet torques to keep the ring particles from spreading past the old boundaries. In either case, the radial spreading of the entire ring system would provide the ultimate release of orbital energy needed to sustain a non-zero level of random velocities[18]. In contrast, if embedded and/or boundary moonlets are constrained by resonant locking to outer moons, the maintenance of random velocities can come ultimately at the expense of a very small change in orbital characteristics of the heavy moon[19]. This may invalidate upper limit estimates of c based on lifetimes of Saturn's rings against viscous diffusion[13].

These general comments may bear on the question of the maintenance of Cassini's division. Currently, Cassini's division and B ring material give a net outward push on the proposed moonlet ('guardian') embedded within the inner rift with a torque per unit mass which can be obtained by substituting $T_j\approx -T_\nu$, with T_ν given by equation (5), into the integral on the right-hand side of equation (9):

$$\frac{\mathrm{d}}{\mathrm{d}t}\{r_j^2\Omega(r_j)\}=-\frac{3\pi}{M_j}[\sigma\nu r^2\Omega]_{\text{B ring}}^{\text{Cassini division}}+T_j^{\text{ext}} \qquad (15)$$

where the quantity in the brackets is to be evaluated well inside the Cassini division and the B ring. If the guardian is unconstrained, it could not prevent the closing of the Cassini division until the ring material reached surface densities comparable with the B ring and could exert an equal but opposite torque to balance that of the B ring. We believe that this compression is prevented by locking of the guardian in the classical 2:1 resonance with Mimas, which is itself locked to Tethys[19]. Using estimates of the guardian moonlet's mass, the viscosity and surface density of the B ring, we find that the Mimas 2:1 lock may barely have enough strength T_j^{ext} to keep the guardian in place. (Relevant formulae can be found in ref. (17).) Unfortunately, the proposed lock feeds eccentricity into Mimas' orbit, and unless there is a means to damp the eccentricity growth, the small current value of Mimas' eccentricity could pose a severe difficulty for this proposed mechanism.

The outer rift moonlet lies between two regions of roughly equal surface density; therefore it may remain stationary even if it is not locked to a heavy outer moon. (Note that the outer rift is located within current uncertainties at the 5:4 resonance with 1980 S27.) The outer edge of Cassini's division is a mystery. It is unlikely that a moonlet maintains the sharp density discontinuity between the division and the A ring. First, no gap is observed here down to the resolution limit of 10 km, so any moonlet present would have to be quite small. Second, the strongest resonance in this region, the 5:4 resonance with 1980 S26, has neither the right sign (for stability) nor sufficient magnitude to hold a small moonlet against the inward push of the A ring.

Independently of the mechanisms operating beyond the rift moonlets, we may use the observed locations of the gaps to fix the moonlet positions r_j at the present epoch. We can then adjust the moonlet masses and the amount of ring matter between pairs of moonlets (M_j/c and C_0 in equation (13)) to give optical depth profiles that best resemble the major structures of the inner Cassini division. An internal check is available as the resulting structure yields, as it should, no appreciable moonlet motions if the rift guardians are held fixed. The best fit using five moonlets is given in Table 1. The resulting theoretical optical depth profile with equal heights of the plateaux (Fig. 2) is a natural consequence of the adopted physics and not of any adjustable parameter of the model. The observed approximately equal spacings of the moonlets may imply nearly uniform conditions in which the moonlets were initially formed.

Discussion

Embedded moonlets may explain why the inner rift of Cassini's division is located at the Mimas 2:1 resonance, and how large optical depth variations can be maintained between the division and the B ring. They also explain Encke's gap, which has sharp edges and no correspondence with resonances of known moons, as well as many of the gaps in the C and D rings. We conjecture that the many ringlets and grooved patterns in the B ring result from bodies massive enough to exert appreciable gravitational torques, but too small to clear resolvable empty gaps. Finally, Voyager 2 observations may detect rift moonlets if they have the sizes listed in Table 1. Gap moonlets could be harder to detect.

We thank Joe Burns, Peter Goldreich, Steve Lubow, Jim Pollack and Bob Sills for discussions. This work was supported in part by NASA consortium grant NC A2-OR050-102.

Received 4 June; accepted 28 July 1981.

1. Smith, B. A. *et al. Science* **212**, 163–191 (1981).
2. Collins, S. A. *et al. Nature* **288**, 439–442 (1980).
3. Cuzzi, J. N., Lissauer, J. J. & Shu, F. H. *Nature* **292**, 703–707 (1981).
4. Tyler, G. L. *et al. Science* **212**, 201–206 (1981).
5. Henon, M. *Nature* (in the press).
6. Jeffreys, H. *Mon. Not. R. astr. Soc.* **107**, 260–262 (1947).

7. Goldreich, P. & Tremaine, S. *Astrophys. J.* **223**, 857–871 (1979).
8. Goldreich, P. & Tremaine, S. *Astrophys. J.* **241**, 425–441 (1980).
9. Lin, D. N. C. & Papaloizou, J. *Mon. Not. R. astr. Soc.* **186**, 799–812 (1979).
10. Elliott, J. L., Dunham, E. & Mink, D. *Nature* **267**, 328–330 (1977).
11. Goldreich, P. & Tremaine, S. *Nature* **277**, 97–99 (1979).
12. Cuzzi, J. N., Durisen, R. H., Burns, J. A. & Hamill, P. *Icarus* **38**, 54–68 (1979).
13. Goldreich, P. & Tremaine, S. *Icarus* **34**, 227–239 (1978).
14. Goldreich, P. & Tremaine, S. *Icarus* **34**, 240–253 (1978).
15. Lynden-Bell, D. & Pringle, J. E. *Mon. Not. R. astr. Soc.* **168**, 603–637 (1974).
16. Goldreich, P. *Mon. Not. R. astr. Soc.* **130**, 159–181 (1965).
17. Sinclair, A. T. *Mon. Not. R. astr. Soc.* **160**, 169–187 (1972).
18. Brahic, A. *Astr. Astrophys.* **54**, 895–907 (1977).
19. Peale, S. J. *A. Rev. Astr. Astrophys.* **14**, 215–246 (1976).
20. Lin, D. N. C. & Bodenheimer, P. *Astrophys. J. Lett.* **248** (in the press).

Distribution of neutral gas and dust near Saturn

E. C. Sittler Jr*, J. D. Scudder* & H. S. Bridge†

* NASA/Goddard Space Flight Center, Laboratory for Extraterrestrial Physics, Greenbelt, Maryland 20771, USA
† Center for Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The distribution of neutral gas and dust within the magnetosphere of Saturn has been inferred from the electron velocity distribution functions measured by the Voyager 1 plasma science experiment. Substantial enhancements of neutral material near Titan and in the vicinity of Enceladus are found. The E ring is also shown to be larger than previously thought.

DURING the Voyager 1 encounter with Saturn, the electron plasma in the energy range 10–5,950 eV was sampled by the plasma science experiment (PLS). The electron distribution function in the vicinity of Saturn is comprised of smoothly joined thermal and suprathermal components; it is not well represented by a single gaussian distribution. We define here the location where the electron distribution departs from this typical behaviour of smooth dependence of phase space density with energy and use this to indicate the presence of neutral material in Saturn's vicinity.

We have reported[1] that in the wake of the satellite Titan anomalous electron distribution functions were sampled by the PLS experiment. Figure 1 shows a representation of the distribution function, $f(v)$, sampled as a function of time as Voyager passed through Titan's wake. The principal anomaly in these spectra is the 'bite-out' region above 700 eV where $f(v)$ is abruptly reduced to values below instrumental threshold. This energy-dependent reduction should be contrasted with the smooth and otherwise regular spectra of the thermal/suprathermal type seen on either side of the bite-out feature. This 'quenching' of the suprathermal part of the spectrum has been interpreted[1] as a result of inelastic collisions of electrons with Titan's extended neutral atmosphere. The extinction becomes strong when the electron gyroradius becomes comparable to, or greater than, the local scale height of the neutral atmosphere[1]. A corresponding enhancement of the low-energy population is observed, which is theoretically expected when the primary electrons become degraded and produce secondaries, and/or when the neutral gas is photoionized.

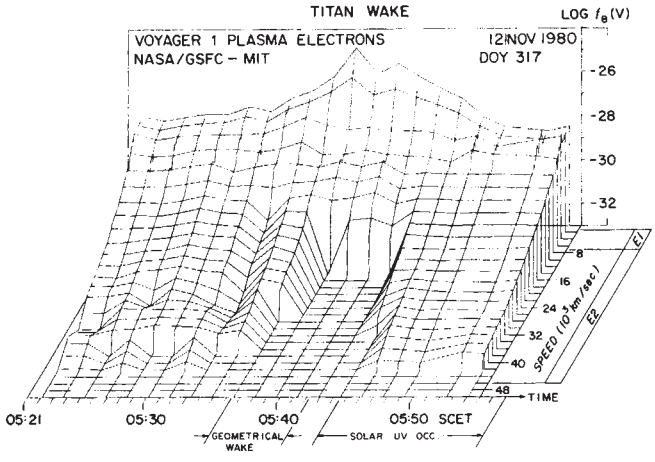


Fig. 1 The observed electron distribution functions in the Titan wake region. Energy increases towards the viewer. An individual distribution is plotted at a time corresponding to measurement of its lower energy portion $E1$. The low and high electron energy modes of the instrument $E1$ and $E2$ are indicated, respectively, where energy mode $E2$ is sampled 48 s after $E1$.

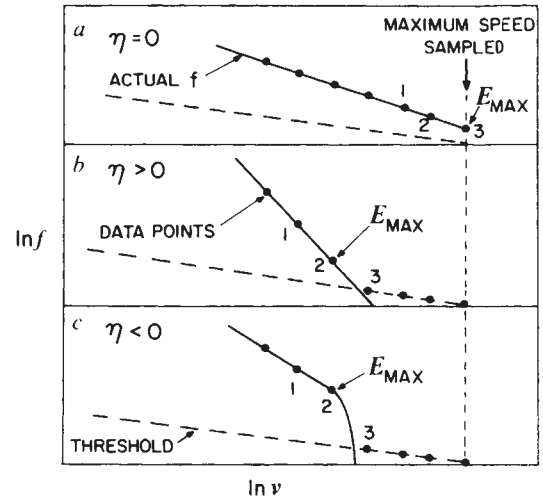


Fig. 2 Log of the distribution function plotted against the logarithm of the speed for three possible values of the curvature parameter (a, b, c). The minimum flux level or threshold and maximum speed sampled by the instrument are indicated. The data points indicated by dots represent the observed distribution function while the actual distribution function being measured by the instrument is displayed by the solid curves. The speed channel for computing E_{max} for each case is denoted. In computing η ponts 1, 2, and 3 are used. To remove any artificial scatter in η when flux levelsare low, the data points 1 and 2 must be five times or more above instrument noise. If not, E_{max} is reduced until this condition is satisfied.

We have systematically located analogous bite-out features in the distribution function samples acquired during encounter. The easiest way to quantify these features is to evaluate the curvature of the distribution function at the maximum energy (E_{max}) that the distribution is observed above the PLS instrument's flux threshold. Here we evaluate a dimensionless curvature indicator, η, given by

$$\eta = \left.\frac{d^2 \ln f(v)}{d(\ln v)^2}\right|_{E_{max}}$$

the second derivative of the logarithm of the observed distribution with respect to the logarithm of the speed, centred at the last energy channel E_{max} which has detectable fluxes above the instrument noise and telemetry thresholds; η vanishes when the local spectrum near E_{max} is a power law in particle speed. This is usual in the magnetosphere in our energy range. Further, the curvature index has been constructed to differentiate between an intrinsic spectral change near threshold ($\eta < 0$) and a spectrum-preserving reduction in the particle flux ($\eta \geq 0$).

Figure 2 shows three possible transitions of the observed speed distribution from above to below instrument threshold. A power law is assumed for the distribution function $f(v)$ where different spectral slopes in Fig. 2a, b, c are used only for illustration. In Fig. 2a the spectrum remains above threshold for the

0028–0836/81/340711—04$01.00
© 1981 Macmillan Journals Ltd

full energy range of the instrument: here $\eta = 0$. Figure 2*b* illustrates the situation where there is a spectral preserving reduction in the actual distribution function; here, because of the instrument threshold level, $\eta > 0$. Figure 2*c* illustrates the negative curvature situation, similar to that shown in Fig. 1 for Titan's wake. Here the spectral trend below E_{max} is shallower than the trend suggested by the fluxes above E_{max}. Because the threshold level is greater than or equal to the incident particle flux above E_{max}, an upper limit for η is determined using the threshold value as the third point in the numerical evaluation of the second derivative. The curvature parameter computed near threshold gives a qualitative measure of the curvature in $f(v)$ at E_{max}; and is used to identify regions where bite-out features are present and to give general trends about the energy dependence of the reduction experienced by the suprathermal electrons. Hence $\eta \geq 0$ indicates no bite-out signature, whereas $\eta < 0$ signifies a bite-out. The more negative the curvature parameter the greater the reduction in $f(v)$.

Observations

We have processed all available electron data acquired close to and inside Saturn's magnetosphere. The results are shown in Fig. 3 with abscissa the equatorial radius, ρ, of the observer in a cylindrical coordinate system with its z axis aligned with Saturn's rotation axis. Within the plot of the L shell parameter the annotated arrows show the equatorial distance, ρ, of field lines which simultaneously thread the spacecraft and various inner satellites according to a model of the magnetic field which incorporates both an intrinsic potential model of Saturn's magnetic field and the first-order effects of a distributed ring current outside 8.7 R_S (Saturn radii)[3]. Positions of the bow shock and magnetopause crossings[1] are also indicated, so changes in electron distribution function can be discussed in relation to the boundaries of the magnetospheric system. Note that the density scale is different for inbound and outbound passes.

In the magnetosphere the curvature parameter is usually near zero whenever E_{max} is at the upper speed channel of the instrument; this supports our interpretation that the parent distribution for the suprathermal electrons is a power law within Saturn's magnetosphere. However, in many places the η-curvature parameter is noticeably and significantly negative. The bite-out signature displayed in the three-dimensional plot of the Titan feature is retrieved as indicated by the sharp reduction in E_{max} and the accompanying increase in the magnitude of the negative curvature. On either side of the principal Titan feature, at least two and possibly more subsidiary density enhancements previously called 'plumes'[1] comprised of electrons cooler than the surrounding magnetospheric electrons can be seen (19.1 and 21 R_S). These density enhancements are associated with changes in the curvature index to more negative values, but at larger values of E_{max} than that at Titan's wake. According to the 'plume' model[1], the density enhancement at 21 R_S should be of most recent origin, because it has a higher density and colder temperature, presumably due to recent contact with the Titan atmosphere. Note that E_{max} for this plume is smaller than that for the other plume seen at 19 R_S, which should be older. In addition to these relatively clear instances of remnants of neutral plasma interactions, there is evidence for bite-out signatures in the suprathermal electrons outside of Titan's orbit all the way to the magnetopause boundary and possibly even into the magnetosheath. Centrifugal forces resulting from co-rotation would tend to move plasma which has interacted with Titan towards the magnetopause. Bite-out signatures observed in the magnetosheath could therefore result from an open magnetic topology, where the suprathermal electrons, which have encountered Titan in the past, have leaked out of the magnetosphere into the magnetosheath. Note that the intrinsic magnetosheath electron spectrum may not be a true power law, but rather maxwellian, in which case the normal curvature of the distribution function crossing the instrument threshold would be

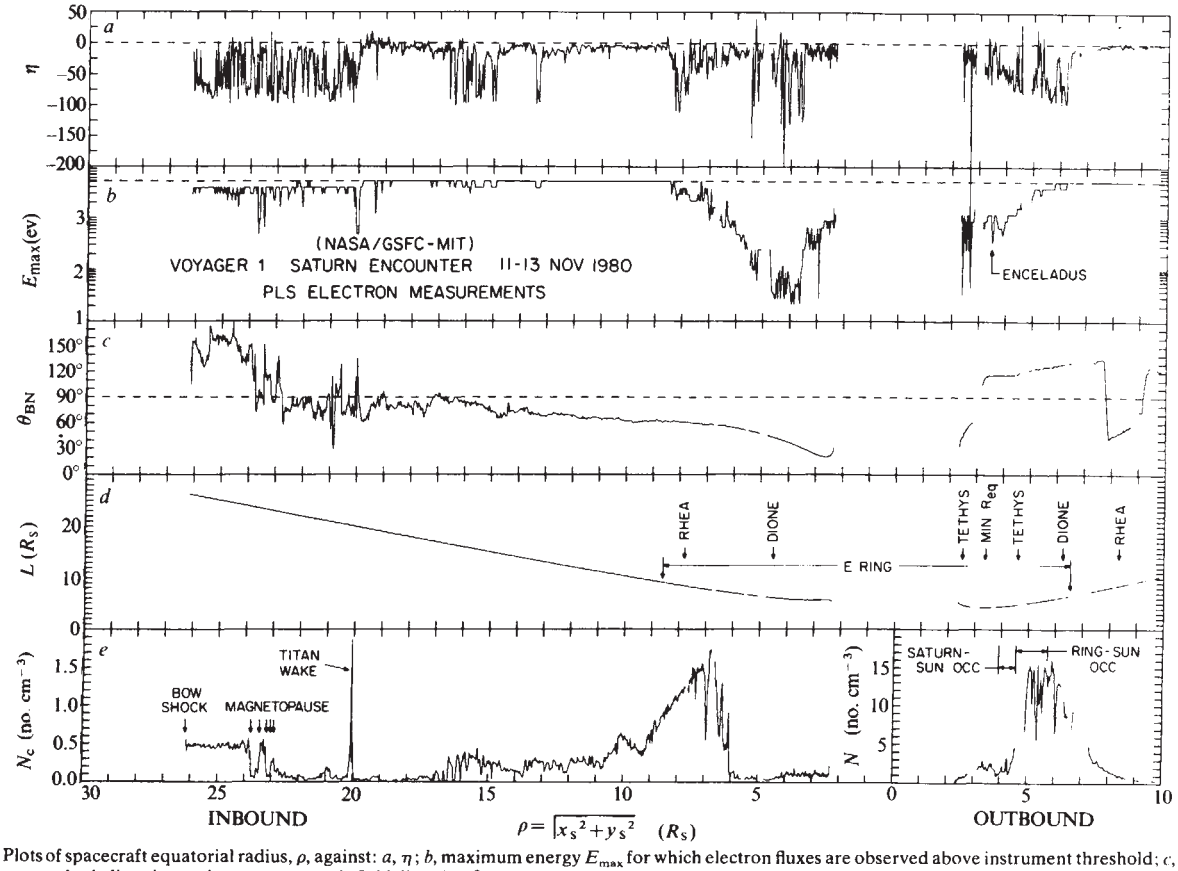


Fig. 3 Plots of spacecraft equatorial radius, ρ, against: *a*, η; *b*, maximum energy E_{max} for which electron fluxes are observed above instrument threshold; *c*, angle θ_{BN} between sensor look direction and current magnetic field direction [2]; *d*, the dipole L shell parameter of the spacecraft; *e*, the electron density for the thermal electrons (low energy component). The Saturn–Sun occultation and ring–sun occultation periods are indicated. The density estimates for these periods show the trend of the data. Complications in estimating the spacecraft floating potential when solar occultation occurs means the analysis is incomplete. Any variations in the density during this period could be artificially produced by unaccounted for variations in the spacecraft floating potential and should be ignored. The spacecraft positions are computed using the Connerney *et al.* model[3] and not from the dipole L shell parameter.

© 1981 Nature Publishing Group

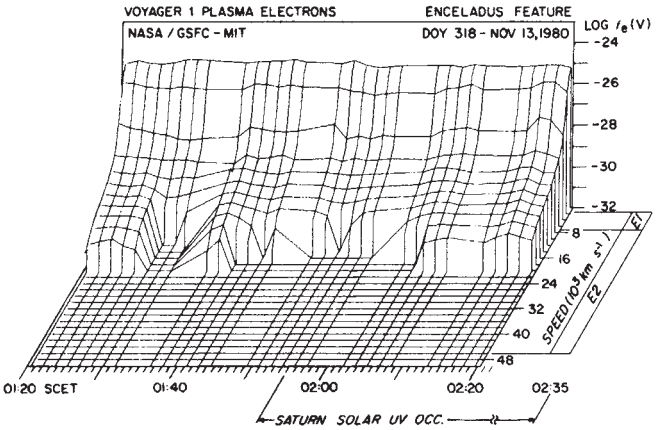


Fig. 4 Plot of the observed electron distribution function for the bite-out event observed near closest approach with Enceladus's L shell (see Fig. 1). When the spacecraft is in shadow, it may reach negative potentials of ~ -10 V significantly below the E_{max} energies reported here. Therefore, only minor trajectory perturbations can be induced by these size potentials for electrons with energies near E_{max}. Hence little effect on the intrinsic bite-out signature is expected when the spacecraft goes into shadow in this region. Data gaps that are ~5 min are centred at 01:36, 01:48, 02:02, and 02:11.

negative. If true, negative curvature signatures for η in the magnetosheath may not be signatures of a bite-out.

Inside Titan's orbit on the inbound traversal between 15 and 17 R_S and briefly at 13.4 R_S there are additional intervals of large negative curvature signatures. In these intervals there is no known satellite or suggested ring or dusty material. We have considered the possibility that when Titan is near local midnight, magnetic field lines which have had close encounters with Titan could map into these smaller radial distances at local noontime. The solar wind compresses Saturn's magnetosphere at noon local time moving field lines in closer to Saturn, and distends them at midnight local time moving the field lines out to greater radial distances. The bite-out signatures inside Titan's orbit at noon local time may then be vestiges of plasma that have experienced a close encounter with Titan near midnight local time. The magnetopause of Saturn has been observed[4,5] as close as 17.3 R_S to Saturn by Pioneer 11 which supports this interpretation. This interpretation also requires a persistence (lifetime) of the curvature signature over a period that is comparable to one-half the orbital period of Titan (8 days or 18 Saturn rotation periods). Losses from these field lines would be smaller than those loaded with plasma by Titan at local noon, as the latter field lines should come much closer to the magnetopause surface, and be extended further down the tail at local midnight where the field lines can open up and lose plasma[6,7].

E ring signature

The curvature parameter indicates another region of suprathermal extinction between 8.4 R_S inbound (317:16:56) and about 6.7 R_S outbound (318:04:45). At the extremities of this interval the curvature parameter is suddenly sharply increased in magnitude, and while remaining negative then begins a recovery towards smaller magnitude but still negative values. The value of E_{max} also exhibits a systematic variation within this interval, decreasing as we proceed towards closest approach and then recovering on the outbound traverse. The amplitude of this variation is remarkable with a range of 6,000–30 eV. Note that the onset of the sharp inbound bite-out signature occurs while the density is still rising so the decrease in the E_{max} parameter cannot result from decreased fluxes without a corresponding spectral change. When the sharp bite-out commences at 8.4 R_S inbound the maximum energy at which particle fluxes are observed above threshold is higher than it is in the weakest curvature parts of this interval. At the outbound termination of the bite-out period at 6.5 R_S, a similar phenomenon is observed: the erosion of $f(v)$ is the most severe as E_{max} rises back towards the maximum PLS energy.

We suggest these signatures are produced by the interaction of particulate matter with the local thermal plasma. We associate our observations with the optically defined E ring first imaged by Feibleman[8] and whose impact on the trapped particles was theoretically predicted by Thomsen and Van Allen[9]. Thomsen and Van Allen showed that the extinction of low energy electrons ($E < 6$ keV) becomes more pronounced as the particle energy is raised. This is essentially the effect we have observed in the relation between the curvature index and E_{max}. The strong energy dependence arises either because the degradation of electrons is proportional to the bounce frequency which is in turn proportional to the particle speed for particles with magnetic mirror points above or below the E ring or, when this is not the case, the collision frequency is proportional to the particle speed (geometrical cross-section).

Thomsen and Van Allen make a prediction about the angular distributions expected for particles interacting with a thin ring of dust, which has been essentially confirmed by Pioneer observations[10,11]. Particles with equatorial pitch angles near 90° are preferentially removed as their path lengths through the ring material are larger and bounce periods shorter than those whose pitch angles are near 0°. Throughout most of the interval given by the E ring in Fig. 3, the PLS electron detector had a mean field of view including particles having 90° pitch angles. For the present energy range and the pitch angles sampled, using known ranges of optical opacities of the E ring[12] and measured diffusion coefficients within Saturn's inner magnetosphere[13], the attenuation of the speed distribution is probably exponentially dependent on the electron speed. This explains the sensitivity of the curvature index as E_{max} rises, and why a sharp definition of the outer edge of the E ring results from this analysis. The PLS definition of the E ring corresponds closely to the definition of the inner boundary of the ring current by the MAG team[3,14]. The energies of electrons used here to define the E ring are substantially below that used by the Pioneer investigations; so they are more sensitive to smaller column densities than previous studies since the stopping distances or ranges for charged particles through neutral material increase with particle energy. The extension of the E ring to these radial distances also explains why the distributed currents of the outer magnetosphere seem to stop in this vicinity. Energetic ions which have ranges or stopping distances less than that of a 6 keV electron are probably the principal current carriers of the azimuthal ring current. These current-carrying ions are then removed by the E ring which determines the inner edge of the ring current.

The radial extent of the E ring is not symmetrical about Saturn; the ring extends to nearly $L = 9.0\ R_S$ in the noon hemisphere and extends to $L = 6.5\ R_S$ in the outbound traverse. A similar asymmetry has been observed[10,15] by Pioneer 11, suggesting this is a permanent feature of the E ring plasma interaction—but no explanation is available.

Within the broad E ring region another feature occurs which is not consistent with the overall trend of the bite-out signatures of the E ring itself. Between 3.7 R_S (day 317 at 21:28 LT) inbound and nearly 2.8 R_S (day 318 at 00:34 LT) outbound there is an anomalous enhancement of the E_{max} parameter. During this interval the magnetic field vector becomes increasingly aligned with the sensor look direction. This change in the relative viewing angle could be reconciled with the observation of the field aligned portion of the theoretically expected dumbbell pitch angle distribution[7], which is outside the field of view of the instrument during the earlier portion of the E ring measurements. Alternatively, this region could be a new particle regime, with a different set of forces and history for the observed particles. This view relies on the results of Acuña *et al.*[14] who report that during this anomalous period a 'non-potential' feature in the magnetic field indicates a distinct spatially localized disruption of the otherwise dipolar magnetic topology. During this period the last channel above threshold increases from 50 eV to nearly 1 keV in near coincidence with the non-potential field perturbation. This shift may represent an energization of the particles or simply an enhancement of their number

© 1981 Nature Publishing Group

density. As the magnetic anomaly requires a field-aligned current and as the high mobility of the electrons argues that they would be the principal carriers, the enhanced electron fluxes in this plasma anomaly may represent the signatures of field aligned currents. The plasma data by themselves cannot, however, determine if there is a net current flow.

Enceladus signature

For the brief interval between 01:20 and 02:10 SCET on 13 November, 1980, within the general E ring bite-out, a strong reduction in the suprathermal electrons occurred similar to that shown in Fig. 1 for the Titan wake, and signified by the sharp depression in E_{max} and the curvature parameter between 3.4 and 4 R_S spacecraft equatorial distance. Figure 4 shows the distribution function for this feature. The L shell distance of the strongest depression in E_{max} centred on 01:30 SCET is 4.35 R_S, very near the minimum L shell of 4.34 R_S predicted by Connerney *et al.*[3] As the dipole L parameter plot indicates, L is approximately constant in this interval and it is therefore difficult to determine the equatorial distance to be ascribed to the strongest bite-out. This distance is close to, but outside, the radial distance of Enceladus, which has been suggested by Terrile and Tokunoga[16] to be the source of the E ring. Terrile and Tokunoga observed a brightening of the E ring in the vicinity of Enceladus. If this interpretation is correct, we have observed bite-out signatures on flux tubes that have traversed this enhancement of dust about Enceladus. At the energies considered here, the previous explanation of the bite-out at Titan is inappropriate because the flux tubes do not come within a gyroradius (<2 km) of Enceladus or any other satellite and the scale lengths for these features are of the order of a Saturn radius. The possibility of an interaction with a localized dense cloud of gas with neutral column densities $\sim 10^{12}$ cm^{-2} cannot be ruled out by these measurements, but the lack of evidence for such a dense cloud of gas from UV observations[17], makes this explanation unlikely.

Conclusions

Based on our suggestion that neutral absorbing material in the vicinity of Saturn can be sensed by studying the spectral modifications to the speed distribution of low energy electrons ($E < 6$ keV) we have remotely identified the neutral gas around Titan which is a strong quencher of the (otherwise commonplace) suprathermal electrons. A similar feature has been used to delineate the E ring, which is considerably larger than previously thought. This results from the use of lower energy electrons as the probes of dust in the ring, as they are stopped by smaller dimension particles. In our energy range the E ring is also asymmetric in local time. The present definition of the extremities of the E ring correspond closely to the inner edge of the ring current as modelled by the MAG team, which argues indirectly that ions with ranges through matter (dust) less than that of a 6 keV electron are the prominent ring current carriers.

The progressive reduction of the suprathermal electrons with decreasing radial distance within the E ring strongly suggests that the suprathermal electrons are diffusing inwards (source must be outside of E ring), whereas outside Titan's orbit we found evidence for outwards diffusion.

The removal of the suprathermal electrons within the E ring explains the apparent collapse of the plasma sheet[1]. Due to extinction of the suprathermal electrons, which dominate the electron pressure outside the E ring, the mean energy of the electrons decreases by a factor of ~5. Consequently, the ability of electrons to leave the equatorial plane and pull off the ions by the polarization electric field, is reduced. The scale height of the density variation should then be smaller within the E ring regime than outside. Estimates of the scale heights within and outside the E ring assuming a composition of singly ionized nitrogen are in good agreement with those inferred by Bridge *et al.*[1]

We interpret a prominent extinction feature in the vicinity of Enceladus to be the result of enhanced dust around Enceladus; this may support earlier speculations that Enceladus[16] is the primary source of the E ring material.

Finally, the electron fluxes are strongly enhanced relative to the ambient conditions within the non-potential magnetic field disturbance within the E ring traversal. These combined observations suggest a localized field aligned current system with the electrons being the principal current carriers.

We acknowledge the advance availability of the magnetic field data from the Voyager magnetometer team and N. F. Ness, and magnetic coordinates provided by the magnetic modelling effort of the MAG group. We thank K. W. Ogilvie and N. F. Ness for helpful comments and M. C. Harrison and L. J. Moriarty for technical assistance.

Received 13 May; accepted 15 July 1981.

1. Bridge, H. S. *et al. Science* **212**, 217–224 (1981).
2. Ness, N. F. *et al. Science* **212**, 211–216 (1981).
3. Connerney, J. E. P., Acuña, M. H. & Ness, N. F. *Nature* **292**, 724–726 (1981).
4. Wolfe, J. H. *et al. Science* **207**, 403–407 (1980).
5. Smith, E. J. *et al. Science* **207**, 407–410 (1980).
6. Hill, T. W., Dessler, A. J. & Michel, F. C. *Geophys. Res. Lett.* **1**, 3 (1974).
7. Carbary, J. F., Hill, T. W. & Dessler, A. J., *J. geophys. Res.* **81**, 5189 (1976).
8. Feibelman, W. A. *Nature* **214**, 793–794 (1967).
9. Thomsen, M. F. & Van Allen, J. A. *Geophys. Res. Lett.* **6**, 893–896 (1979).
10. Simpson, J. A. *et al. J. geophys. Res.* **85**, 5731–5762 (1980).
11. Bastian, T. S., Chenette, D. L. & Simpson, J. A. *J. geophys. Res.* **85**, 5731–5762 (1980).
12. Smith, B. A. *The Saturn System* 105–111 (NASA Conference Publication 2068, 1978).
13. Van Allen, J. A. Thomsen, M. F. & Randall, B. A. *J. geophys. Res.* **85**, 5709–5718 (1980).
14. Acuña, M. H., Connerney, J. E. P. & Ness, N. F. *Nature* **292**, 721–724 (1981).
15. McDonald, F. B., Schardt, A. W. & Trainor, J. H. *J. geophys. Res.* **85**, 5813–5830 (1980).
16. Terrile, R. J. & Tokunaga, A. *Bull. Am astr. Soc.* **12**, 701 (1980).
17. Broadfoot, A. L. *et al. Science* **212**, 206–211 (1981).

Low-frequency plasma waves near Saturn

B. M. Pedersen*, M. G. Aubier*, & J. K. Alexander†

* Observatoire de Paris, Section d'Astrophysique de Meudon, 92190 Meudon, France

† Laboratory for Extraterrestrial Physics, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Voyager planetary radio astronomy observations of low frequency emissions detected around the time of closest approach to Saturn and near the outbound ring plane crossing are presented. Near the ring plane an electron density of between 5 and 20 electrons cm^{-3} at distances of $\sim 6R_S$ is estimated.

THE low-frequency plasma wave emissions presented here are reminiscent of emissions observed within the Earth's magnetosphere[1] and more recently within the jovian magnetosphere[2,3]. Their most striking feature consists of intense spiky bursts, interpreted as local electrostatic upper hybrid resonances accompanied by odd-half-harmonic cyclotron, $(n + 1/2)f_g$ waves where f_g is the electron gyrofrequency. These emissions are preceded by two distinct components at frequencies lower than the normal Saturn kilometric radiation (SKR)[4]: the first one is a polarized drifting emission between 3 and 40 kHz; the second, appearing at still lower frequencies, is unpolarized and narrow banded (1–3 kHz). By analogy with observations in the terrestrial magnetosphere, these two emissions are interpreted respectively as non-thermal continuum radiations[5] and VLF hiss.

The characteristics of the Voyager 1 trajectory close to Saturn are shown in Fig. 1. During the inbound segment, the spacecraft crossed the saturnian equatorial plane downwards at the orbit of Titan. The planetary radio astronomy (PRA)[6] observations during this Titan flyby will be discussed elsewhere[7]. Up to the

0028–0836/81/340714—03$01.00

© 1981 Macmillan Journals Ltd

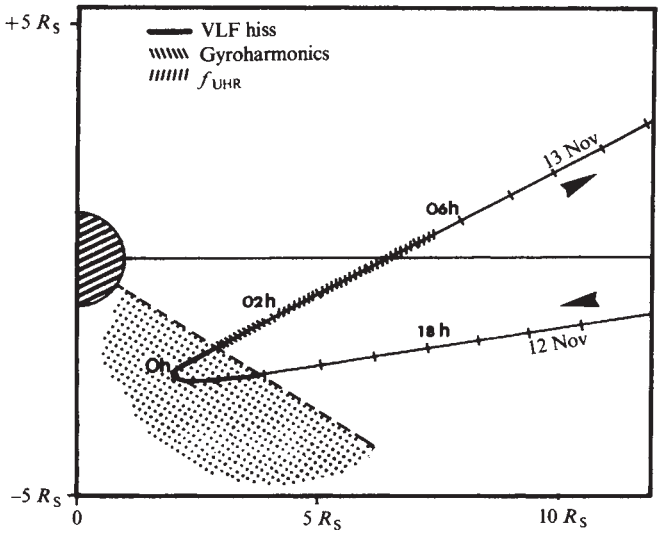


Fig. 1 A meridian plane projection of the Voyager 1 trajectory around closest approach is coded to indicate the locations where three different classes of emission are detected by the PRA experiment. From the f_{UHR} resonance we deduce electron densities between 5 and 20 electrons cm^{-3} (see Fig. 3). The dashed line corresponds to magnetic latitude $ML = -28°$.

time of the closest approach, the spacecraft continued towards higher negative saturnian latitudes (also magnetic latitudes), while moving around the planet. During the outbound pass through the saturnian magnetosphere, Voyager 1 returned towards the ring plane, which it crossed at the orbit of the moon Dione ($6R_S$) and then continued towards high positive latitudes.

The different low-frequency emissions observed near periapsis are shown as a function of time on the fixed frequency plots in Fig. 2. Each channel has a 1-kHz bandwidth, and each point represents a 6-s measurement interval. Two traces in a particular frequency channel correspond to signal levels in each of the opposite senses of circular polarization and indicate that the emission is polarized.

The emission appearing at frequencies above ~60 kHz in Fig. 2 is freely propagating SKR[8], whereas the broad-banded emission present in all channels around the time of the outbound ring plane crossing does not show normal SKR characteristics. In particular, it does not appear at the typical saturnian longitudes for the low-frequency SKR component. We cannot explain this emission which is confined to a region around the equatorial plane. We now consider the activity which occurs only in the frequency range below 60 kHz and which can provide information about plasma conditions in the vicinity of the spacecraft.

Impulsive emissions appear between ~1 h and 6 h in the three lowest-frequency channels. The very intense emission in the 1.2-kHz channel which saturates our receiver most of the time (intensity level > 10^{-11} $V^2 m^{-2} Hz^{-1}$) appears below half the local electron gyrofrequency and can be thus interpreted as VLF noise[9]. Another type of emission in this time interval consists of brief spiky bursts (Fig. 2). This impulsive emission first develops at low frequencies, moves towards high frequencies and then back to low frequencies. We interpret this pattern as a variation of the local upper hybrid resonance frequency along the spacecraft trajectory. This interpretation is supported by the resemblance to similar narrow-banded emissions observed in the magnetospheres of Earth[1] and Jupiter[2,3]: The electrostatic noise at $f_{UHR} = (f_p^2 + f_g^2)^{1/2}$, where f_p is the local plasma frequency, allows a direct estimate of the electron density at the position of the spacecraft. The resulting density curve around closest approach is shown in Fig. 3 together with density curves estimated by Voyager 1 plasma wave (PWS)[9] and plasma science experiments (PLS)[10]. A time discrepancy is noted with respect to PWS results, whereas our maximum density (~20 electrons cm^{-3}) is only slightly higher than the PLS value.

Preceding this developing f_{UHR} line we observe smooth emission humps of longer duration on the 20.4 and 39.6 kHz

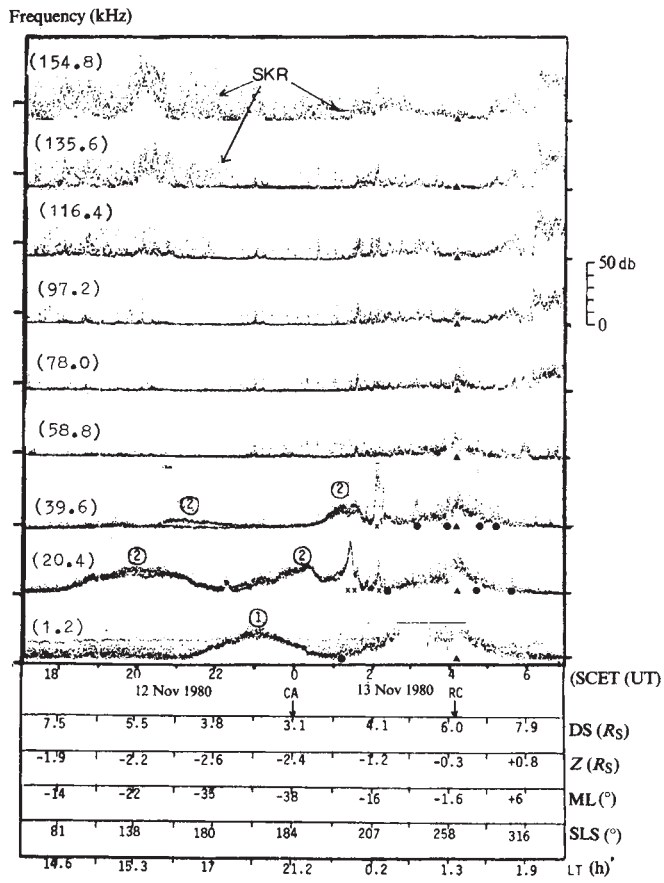


Fig. 2 Fixed frequency plots (intensity in dB $\mu V^{-1} kHz^{-1}$) of the lowest PRA frequencies as a function of the distance (D_S) from Saturn. The distance (Z) from the equatorial plane, the magnetic latitude (ML), the Subspacecraft Saturn Longitude (SLS) and the local time (LT) are shown. CA denotes the point of closest approach, and RC denotes the ring plane crossing. The broad-banded emission around RC (▲), the $(n+1/2)f_g$ waves (×), the f_{UHR} spiky bursts (●), the VLF hiss ① and escaping ② continuum radiations are indicated. The intense spiky emission from 2 h to 5 h on the 1.2-kHz channel is interpreted as VLF noise. The emissions observed throughout the period at frequencies ≥60 kHz are due to freely propagating Saturn kilometric radiation.

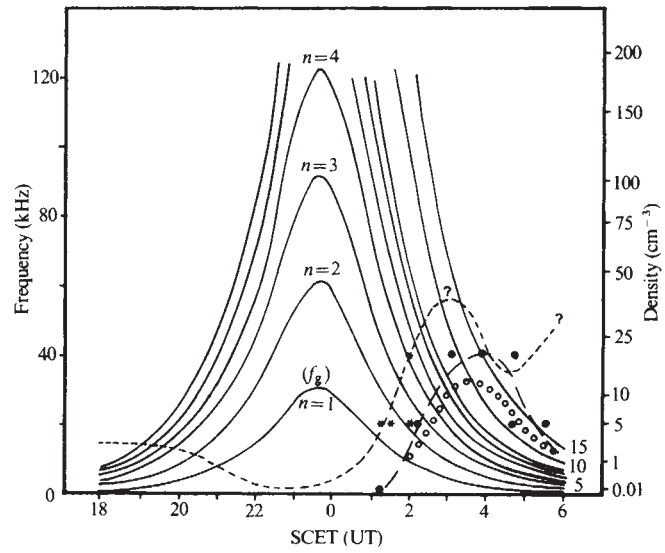


Fig. 3 Appearance of electrostatic emissions (×, smooth $(n+1/2)f_g$ humps; ●, f_{UHR} spiky resonances) with respect to the evolution of the gyroharmonic pattern as calculated from the measured magnetic field (left ordinate). The density profile deduced from the PRA observations (— — —) is compared with the Voyager PLS[10] (○○○○○○) and PWS[9] (- - - -) results (right ordinate).

© 1981 Nature Publishing Group

channels (Fig. 2). We interpret these emissions as $(n+1/2)f_g$ harmonic waves. Such waves are also observed in the vicinity of the Earth[1] and within the Io plasma torus[2,3] and generally accompany f_{UHR} waves. Figure 3 shows the position of these waves with respect to the gyroharmonic pattern calculated from Voyager 1 magnetic field measurements[11]. The location of the gyroharmonic waves in relation to the f_{UHR} waves can be interpreted in terms of the relative temperatures of a two-component electron distribution consisting of a 'cold' background plasma and a 'hot' component that provides a source of free energy for the waves[3].

In the Earth's and Jupiter's magnetosphere[12] the whole pattern of $(n+1/2)\ f_g$ and f_{UHR} electrostatic waves in some cases seems to be closely associated with two components of non-thermal electromagnetic radiation: an escaping and a trapped continuum component. The emission of type 2 in Fig. 2 resembles such a non-thermal continuum: this polarized emission drifts between frequencies ≤ 20 kHz and ~ 40 kHz and is typical of an apparently distinct narrow-banded component of SKR that is seen intermittently in the PRA data throughout the encounter period[4]. Its characteristics are similar to the narrow-band jovian kilometric radiation (nKOM)[13], and both may originate in the same physical process as the escaping non-thermal continuum radiation[12]. At lower frequencies around 1 kHz a smooth, intense and unpolarized emission 1 (on Fig. 2) is observed during a short period at latitudes lower than $-28°$ (Fig. 1). Gurnett *et al*[9] interpret this low frequency component, below the plasma frequency and far below the gyrofrequency, as VLF noise.

We have shown that the different emission components of the low-frequency spectrum observed during the Saturn encounter constitute a similar pattern to the spectra observed elsewhere (see Fig. 9 of ref. 12). The combination of strong, impulsive bursts at f_{UHR} and smooth emissions at $(n+1/2)f_g$ is common in plasma populations consisting of a mixture of relatively cold and hot energetic magnetospheric plasmas. This mixture, as pointed out for the terrestrial and jovian magnetospheres by Kurth *et al.*[12], agrees with the emission mechanism proposed by Melrose[14], according to which the non-thermal radio continua result from conversion of upper hybrid waves into electromagnetic waves by coalescence with low-frequency waves. Finally our estimate of the electon densities near the time of the outbound ring plane crossing ($4 \leq R \leq 7R_S$) agrees reasonably well with the direct measurements by the PLS team[10].

Received 13 May: accepted 20 July 1981.

1. Christiansen, P. J. *et al. Space Sci. Rev.* **22**, 383–400 (1978).
2. Warwick, J. W. *et al. Science* **204**, 995–998 (1979).
3. Birmingham, T. J., Alexander, J. K., Desch, M. D. Hubbard, R. F. & Pedersen, B. M. *J. geophys. Res.* (in the press).
4. Warwick, J. W. *et al. Science* **212**, 239–243 (1981).
5. Gurnett, D. A. *J. geophys. Res.* **80**, 2751–2763 (1975).
6. Warwick, J. W., Pearce, J. B., Peltzer, R. G. & Riddle, A. C. *Space Sci. Rev.* **21**, 309–327 (1977).
7. Daigne, G., Pedersen, B. M., Desch, M. D. & Kaiser, M. L. *J. geophys. Res.* (submitted).
8. Kaiser, M. L., Desch, M. D. & Lecacheux, A. *Nature* **292**, 731–733 (1981).
9. Gurnett, D. A., Kurth, W. S. & Scarf, F. L. *Science* **212**, 235–239 (1981).
10. Bridge, H. S. *et al. Science* **212**, 217–224 (1981).
11. Ness, N. F. *et al. Science* **212**, 211–217 (1981).
12. Kurth, W. S., Gurnett, D. A. & Anderson, R. R. Preprint 80-42 (University of Iowa 1980).
13. Kaiser, M. L. & Desch, M. D. *Geophys. Res. Lett.* **7**, 389–392 (1980).
14. Melrose, D. B. *J. geophys. Res.* **86**, 30–36 (1981).

Impulsive radio discharges near Saturn

David R. Evans*, James W. Warwick*, Jeffrey B. Pearce*, Thomas D. Carr† & John J. Schauble†

*Radiophysics, Inc, 1885 33rd Street, Boulder, Colorado 80301, USA
†Department of Astronomy, University of Florida, Gainesville, Florida 32611, USA

An unexpected type of emission was observed during the Voyager 1 Saturn encounter. This consisted of periodic episodes of many impulsive discharges throughout the frequency range (20.4 kHz to 40.2 MHz) of the planetary radio astronomy experiment.

NON-THERMAL radio emissions from the Saturn system were first detected by the Voyager planetary radio astronomy (PRA) experiment on board Voyager 1 in January 1980. Since then emission between 100 kHz and ~1 MHz from the planet, termed Saturn kilometric radiation (SKR), has been detected almost continuously. Observations of SKR are detailed elsewhere[1]. For only a few days around encounter a broadband (100 kHz to at least 40 MHz) emission was observed. Initial observations of this form of emission are given elsewhere[2]; here we detail the further analysis of these data.

In its normal mode, the PRA receiver scans downward in frequency through 198 channels between 40 MHz and 1.2 kHz and samples each channel for 30 ms. Each scan includes experiment status information, and requires 6 s (ref. 3). Figure 1 shows part of a typical dynamic spectrum obtained by the receiver in this mode near encounter, in which several vertical streaks are apparent. We believe[2] that these streaks have a physical cause outside, and independent of, the spacecraft. The detailed structure of an individual streak (or 'event') can be determined only on the rare occasions that the PRA receiver is operating in its 'high rate' mode, when two preselected pairs of adjacent frequency channels are sampled simultaneously once every 140 μs. Although we only have 48 s of such data which exhibit clear events, obtained ~4 h after encounter, this is enough to examine the detailed structure of such events, which in many ways resembles the radio signature of a terrestrial lightning discharge. This, and properties previously listed[2], leads us to conclude that we are observing the effects of short-lived electrostatic discharges originating in the vicinity of Saturn (hence Saturn electrostatic discharges (SED)).

To understand better the properties of SED, an algorithm was

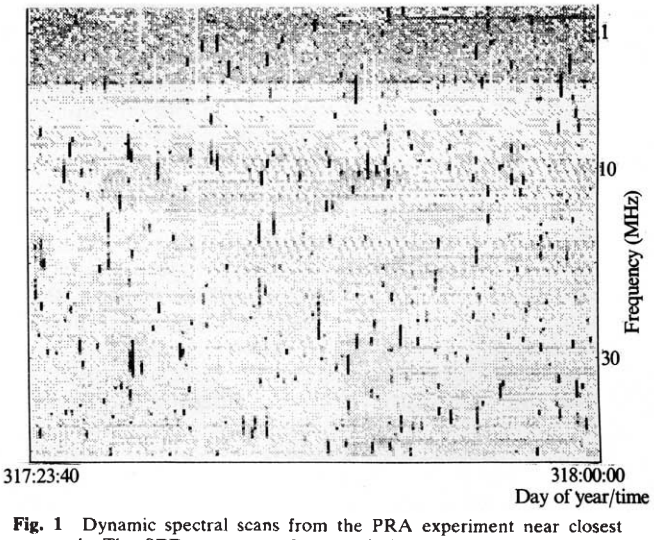


Fig. 1 Dynamic spectral scans from the PRA experiment near closest approach. The SED appear as short vertical streaks whose duration is proportional to the height of the streaks. The broadband nature of the phenomenon is clear.

0028-0836/81/340716—03$01.00

© 1981 Macmillan Journals Ltd

written to identify and log each individual discharge event occurring in the regular low rate data. This proved to be much more sensitive than the plotting of dynamic spectra as in Fig. 1. The rate of discharge occurrence varied markedly throughout the encounter period with the SED being generally confined to apparently periodic episodes. Given the relatively limited data base, and the rapidly changing geometry of the Saturn system as seen from the spacecraft during encounter, the inherent periodicity of these episodes cannot be estimated precisely. We calculate, however, that the observed events are consistent with a source revolving about the planet with a period of ~10 h 10 min (±5 min), radiating away from the planet. The possibility of a corotating phenomenon (of period near 10 h 40 min), previously thought to be a plausible explanation of the observations, is now clearly excluded. Generally, SED episodes are observed over slightly more than 50% of this orbit, which is located about $1.81R_S$ (Saturn radii) from the centre of the planet. Figure 2 shows all the clearly defined episodes we observed; each frame represents one revolution of a reference frame rotating with period 10 h 10 min about the centre of Saturn (0° is defined as that longitude facing the vernal equinox of Saturn at 1.0 January 1970), with spacecraft motion taken into account. The episodic nature of the phenomenon is clear. For convenience, the episodes have been assigned identifying numbers as in Fig. 2.

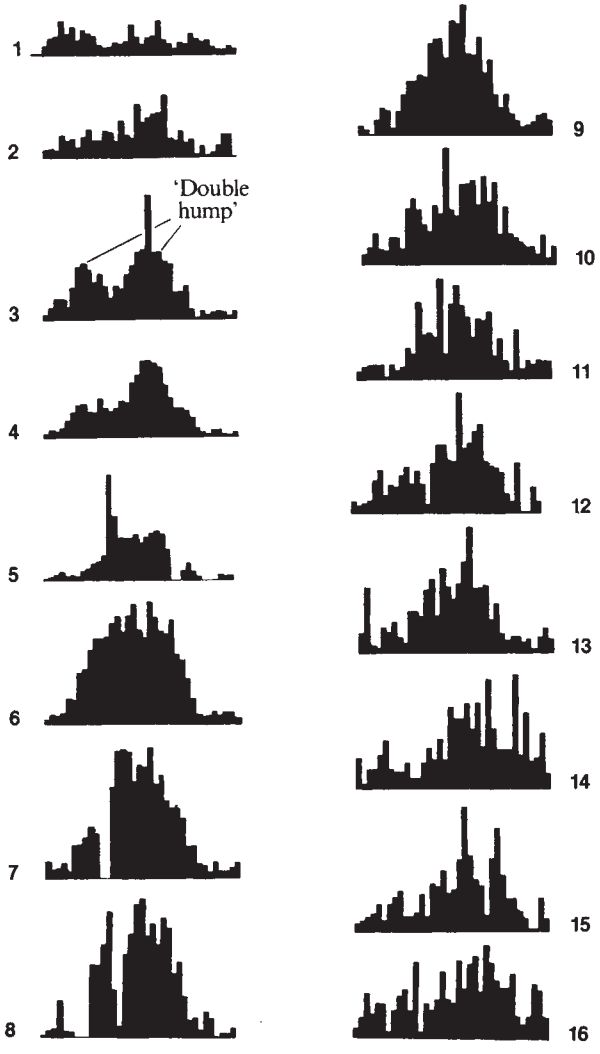


Fig. 2 The episodic gross structure of SED which shows the 16 relatively well-defined episodes (closest approach takes place during episode 5). Each episode consists of 36 bins each 10° longitude wide, representing the sub-spacecraft longitude of a system rotating about Saturn's centre once every 10 h 10 min. The length of each bar is a function of the numbers of observed SED events and the spacecraft distance from $r = 1.81R_S$. This convolution increases the relative noise background with increasing distance from $r = 1.81R_S$.

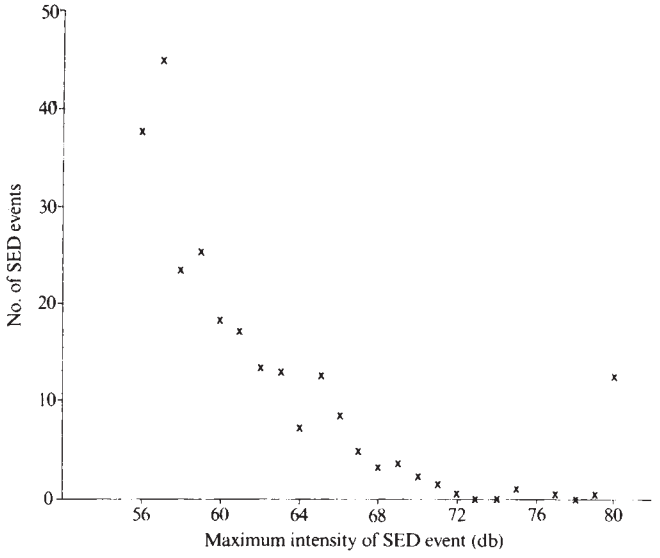


Fig. 3 Maximum intensity of SED bursts for the available high-rate data. Instrument cutoff occurs at +80 dB. The number of points at or above cutoff is highly anomalous and indicates indirectly the possibility of two distinct populations of SED.

Dynamic spectra such as Fig. 1 exhibit clearly the broadband nature of SED. Generally whilst within an episode, we observed SED in all channels which were not swamped by SKR emission. This normally limited us to frequencies ≥800 kHz. For part of the encounter episodes, however, SKR was absent at the lowest frequencies, and SED was clear even below 100 kHz. To test the broadbandedness of SED, spectra consisting of the number of SED events observed per channel on an episode-by-episode basis were calculated. At encounter (episode 5) the spectrum appears remarkably flat; what features are present are due to the sensitivity curve of the receiver, and the changing shape of the spectrum as the spacecraft moves in the days around closest approach indicates that at encounter relatively few discharges go undetected.

Note also that the onset of an SED episode is not uniformly simultaneous at all frequencies. In particular, the encounter episode was observed at 40 MHz some 4.5 h before it was seen at 10 MHz. The start of other episodes and the end of all episodes seem to be independent of frequency.

A marked feature of Fig. 2 is the asymmetry about closest approach; the earliest recognizable SED were observed only ~36 h before closest approach, whereas post-encounter episodes continued for some days. Two of the pre-encounter episodes exhibit a clearly double-humped envelope, a feature

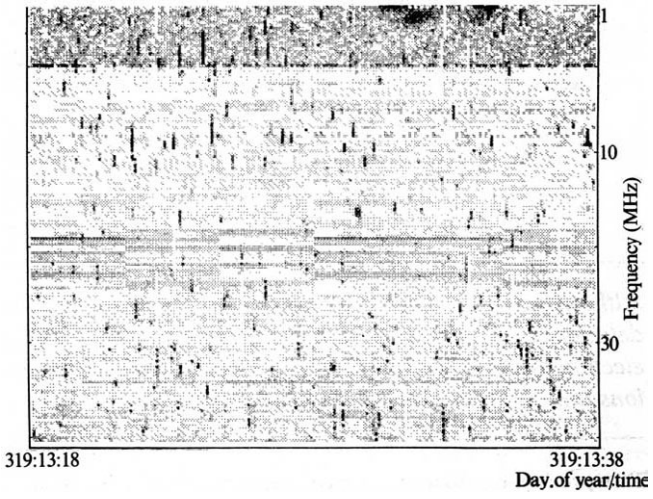


Fig. 4 Dynamic spectrum similar to Fig. 1 for episode 6. The alternating light and dark squares during many SED events indicate the presence of circular polarization, in this case overwhelmingly in the left-hand sense.

©1981Nature Publishing Group

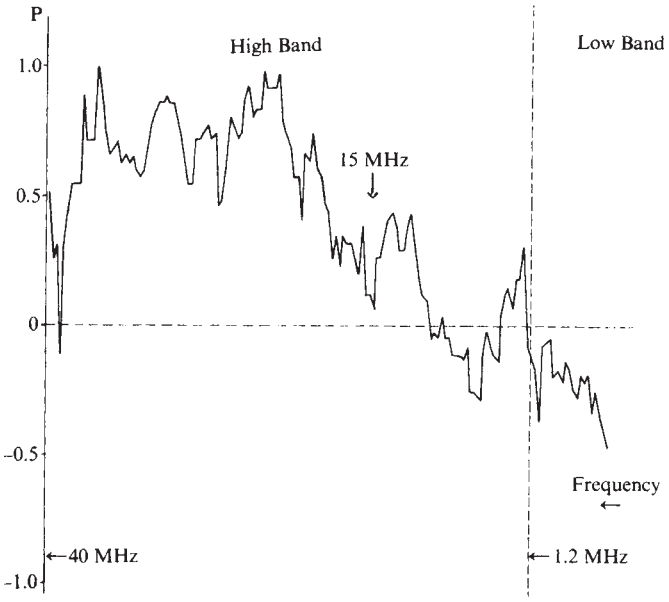


Fig. 5 Polarization as a function of frequency for post-encounter episode. Frequency runs between 40 MHz at left to ~800 kHz at right. A three-point running mean has been used to smooth the curve. Very few points were available at the top of the high band, but at other frequencies above about 15 MHz (marked with arrow) most bursts are clearly left-circularly polarized. Below 15 MHz there is little evidence for any systematic sense of polarization.

which does not appear to be present after the encounter. There is some evidence that this could be due to a smaller source of SED revolving about the planet with a longer period (~10 h 30 min). Pre-encounter, the two sources are just resolved. During the period displayed in Fig. 2 the faster source catches up with, and may overtake, the slower source, and so for the strongest episodes (numbers 3–9), the two are unresolved. In support of this, the early episodes seem to have a slightly greater angular width than the later ones. Support for a two-source model comes from an investigation of the intensity characteristics of the bursts detected in the high rate mode, which exhibit results typical of the superposition of two or more populations (Fig. 3). This hypothesis cannot be rigorously checked until the Voyager 2 encounter.

In the normal mode, the PRA receiver is sensitive to circular polarization. The sense of polarization sampled alternates from channel to channel as a 6-s scan progresses. (Alternating 6-s scans start with alternating polarization sampling of the first channel.) Consequently, should a given SED event show circular polarization, it would appear as a streak of alternating light and dark squares on a dynamic spectrum plot. A search for this phenomenon yielded many examples, such as shown in Fig. 4. We defined an index of polarization for the ith channel, $P(i)$, by

$$P(i) = (L(i) - R(i))/(R(i) + L(i))$$

where $R(i)$, $L(i)$ are the number of 30 ms samples of right-hand and left-hand polarization of the ith channel in which an SED signal was detected. An examination of the SED on the basis of polarization showed that episodes 1–5 contained very few clearly polarized events out of many thousands. Post-encounter episodes, however, were generally strongly left-hand circularly polarized. Polarization/frequency plots were produced for each episode. Figure 5 shows the results for the first post-encounter episode. The SED are clearly strongly left-circularly polarized above ~15 MHz. Below this frequency, the degree of circular polarization is much less, and the sense is frequency-dependent.

We have previously citeh nn on polarization as being evidence of a discharge[2]. Despite the present finding, we still believe that we are observing electrostatic discharges, which are polarized by virtue of either magnetic or curved electric fields, and consequently we have retained the term 'SED'.

In conclusion, some eight characteristics have now been fairly well defined by the Voyager 1 encounter:

(1) a very flat, broadband frequency spectrum;
(2) a period of ~10 h 10 min;
(3) a change in the envelope shape of episodes between pre- and post-encounter;
(4) an intensity population structure typical of plural populations;
(5) episodic structure of width ~180°;
(6) post-encounter episodes continue for about three times as long as pre-encounter ones;
(7) post-encounter bursts are left-circularly polarized at high frequencies;
(8) at least one episode shows the onset of high frequency events some time before that of lower frequency ones.

Any theory for the generation of SED should account for all these characteristics.

We thank other members of the PRA team for helpful suggestions and criticisms of this work. It was supported by NASA-Jet Propulsion Laboratory.

Received 13 May; accepted 15 July 1981.

1. Kaiser, M. L., Desch, M. D. & Lecacheux, A. *Nature* **292**, 731–733 (1981).
2. Warwick, J. W. *et al. Science* **212**, 239–243 (1981).
3. Lang, G. J. & Peltzer, R. G. *IEEE Trans.* **AES-13**, 466–471 (1977).

X-ray and energetic neutral particle emission from Saturn's magnetosphere

E. Kirsch*, S. M. Krimigis†, W. H. Ip* & G. Gloeckler‡

*Max-Planck Institute for Aeronomy, D-3411 Katlenburg-Lindau 3, FRG
† Applied Physics Laboratory, The Johns Hopkins University, Laurel, Maryland 20707, USA
‡ Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742, USA

Although Voyager 1 was not equipped for the detection of X-rays and neutral particles, its low energy charged particle detector (LECP) records suggest a significant flux of these radiations. X-rays could be due to substantial precipitating electron fluxes in the auroral region or the rings whereas energetic neutrals could be due to charge-exchange between trapped ions and Saturn's neutral hydrogen disk.

THE production of auroral X rays and energetic neutral particles is well known in the Earth's magnetosphere (X rays[1]; charge exchange energetic neutral particles[2–5]) and can be expected at Jupiter and Saturn. In the case of the outer planets, besides charge exchange with the planetary exospheric neutrals, the magnetospheric particles are also subject to similar loss processes due to interaction with the neutral clouds emitted from satellites[6].

0028–0836/81/340718—04$01.00
© 1981 Macmillan Journals Ltd

We present here observations of the Voyager 1 spacecraft obtained a few days before encounter of the Saturn magnetosphere in November 1980. As Voyager 1 has no special experiment onboard for X-ray and neutral particle detection, we use the sensitivity and directional measurements of the LECP (low energy charged particle) analyser to identify the neutral radiation above the charged particle background.

Method of detection

The method has been outlined earlier[7] and will be summarized only briefly here. The LECP telescope[8] uses a silicon detector of 96.5 μm thickness and 0.08 cm^2 area to accumulate counts in eight separate sectors near the ecliptic plane. Electrons ≲400 keV are magnetically deflected and cannot reach the detector. A stepwise integration of the X-ray sensitivity of the silicon detector for channel PL01 which has an electronic threshold of 26–31 keV, yields an average sensitivity $\bar{\varepsilon}_1 \approx$ 2.38%/keV, and for channel PL02 $\bar{\varepsilon}_2 \approx 0.8\%$/keV ($E = 31$–60 keV). The detection process for energetic neutral particles is the same as for charged particles: the neutrals become ionized when they hit the detector surface and deposit their energy through ionization energy loss in the detector.

The energy thresholds for neutral or charged particles are:

PL01: 40–53 keV ($\varepsilon \approx 0.4$)

PL02: 53–85 keV ($\varepsilon \approx 1$)

PL03: 85–139 keV ($\varepsilon \approx 1$)

Figure 1 shows the trajectory of Voyager 1 for days 307–330, 1980 and the viewing directions of the eight sectors of the instrument. It can be seen that X rays and energetic neutrals escaping from the magnetosphere should appear in sector 5. Encounter of the bowshock[9] occurred at ~23:26 UT on day 316 (11 November 1980), thus the days before 316 can be used for our study. Criteria for the identification of neutral radiation during the pre-encounter phase (Fig. 1 and Kirsch *et al.* unpublished data) can be summarized as follows:

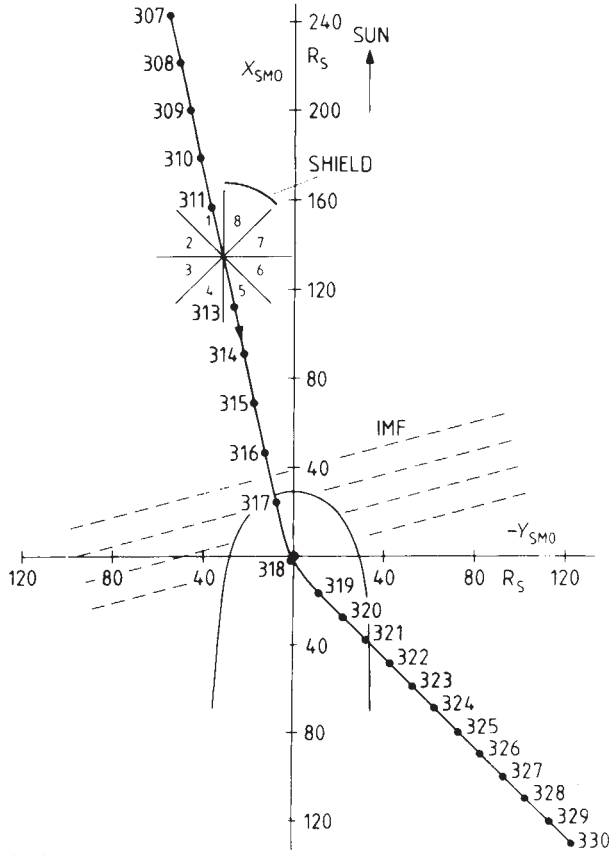


Fig. 1 Projection of the Voyager 1 trajectory on the ecliptic plane (days 307–330, 1980). Indicated are the eight viewing directions of the experiment. Sector 8 is permanently blocked by an aluminum shield.

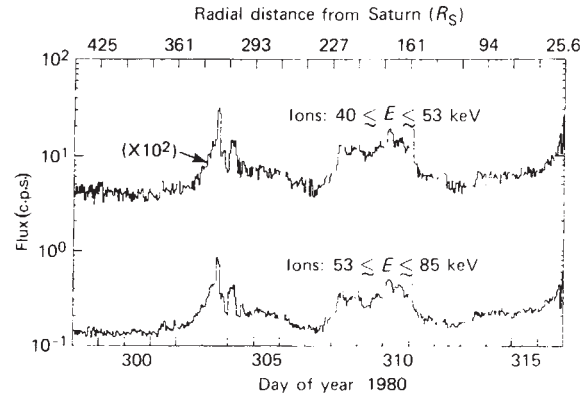


Fig. 2 Scan-averaged measurements of the ion channels PL01 and PL02 for days 298–316, 1980. Interplanetary particle acceleration by Corotating Interaction Regions (CIR) are evident beginning on days 302 and 308.

(1) The interplanetary magnetic field must be nearly perpendicular to the Saturn–Sun line (in the nominal direction) to exclude charged particles escaping from the planet.
(2) The pre-encounter measurements must not be disturbed by solar particles or particles accelerated in corotating structures (CIR) in the interplanetary medium.
(3) Statistically significant enhancements above normal background must appear in the sector which is pointing towards the planet.
(4) The energy spectrum measured within the sector pointing towards the planet should exhibit differences from spectra measured in the other sectors.
(5) The additional flux (after subtracting background) must show a $1/R^2$ dependence (R distance from the planet).

The magnetometer data (ref. 10 and N. F. Ness, personal communication) show the interplanetary magnetic field direction to be almost perpendicular to the Saturn–Sun line, which satisfies criterion (1). To investigate the remaining criteria we show in Fig. 2 the scan-averaged particle measurements for days 298–314. Days 298–300 show the lowest count rates, which are probably due to penetrating galactic cosmic rays. Days 301–305 and 308–311 must be excluded due to the presence of ions associated with typical CIRs (see criterion (2)). On days 313–314 a slight increase was observed which is probably due to solar particles. Thus, we shall examine days 298–300, 306–307, 312–316 (~07:00 UT) in the context of criteria (3)–(5) for the presence of neutral radiation.

Observations

Figure 3*a* shows the sector count rates of channels PL01, 2, 3 for the selected days and the net flux (ΔP) for these channels after the galactic cosmic-ray background was subtracted. Figure 3*b* shows the count rate ratios before, and Fig. 3*c* after, subtraction of the galactic background flux (average of days 298, 299 and 300 has been used as galactic background). From day 312 onward sector 5 shows an increase in comparison with the neighbouring sectors which is ~4 times the statistical error. However, similar peaks exist in sector 1 throughout this period, and in sector 7 from day 313 onwards. Because both of these sectors are generally viewing the solar direction (Fig. 1), we attribute their response to low level fluxes of energetic solar-interplanetary particles. The solar background complicates our detection method. As pointed out in Fig. 1, sector 8 is an experiment calibration sector and cannot be used for further diagnostics of the solar particle contribution. Enhanced fluxes in sector 1 for channels PL02, 3 are most likely produced by the $\vec{E} \times \vec{B}$ drift. The enhanced rate in PL01, sector 1, is due to residual sensitivity to solar X rays, as the Sun was viewed by this sector at this time.

A clue to the identity of the observed fluxes can be obtained by examining the directional energy spectra through the channel ratios. It is evident from the channel ratios presented in Fig. 3*b* (background included) and especially Fig. 3*c* (background subtracted) that the additional flux counted in sector 5 has a

© 1981 Nature Publishing Group

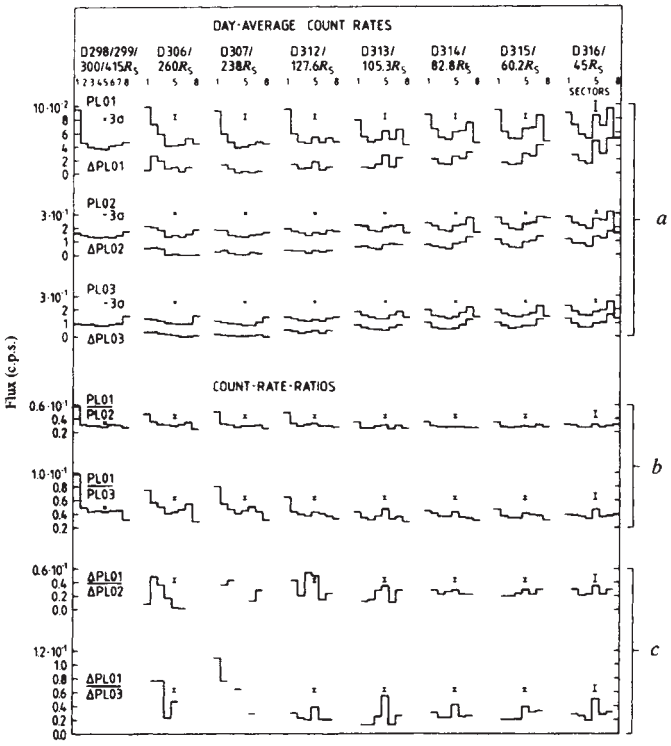


Fig. 3 *a*, Sector-averaged measurements for the days indicated for channels $PL0_{1,2,3}$. $\Delta PL0_{1,2,3}$ indicates that the galactic background is subtracted. The bars on sector 5 represent 3 times the statistical error. *b*, The count rate ratios PL01/PL02, PL01/PL03. *c*, ΔPL01/ΔPL02, ΔPL01/ΔPL03 (after background subtraction) show that sector 5 measured an additional flux which has a softer spectrum than the flux in neighbouring sectors.

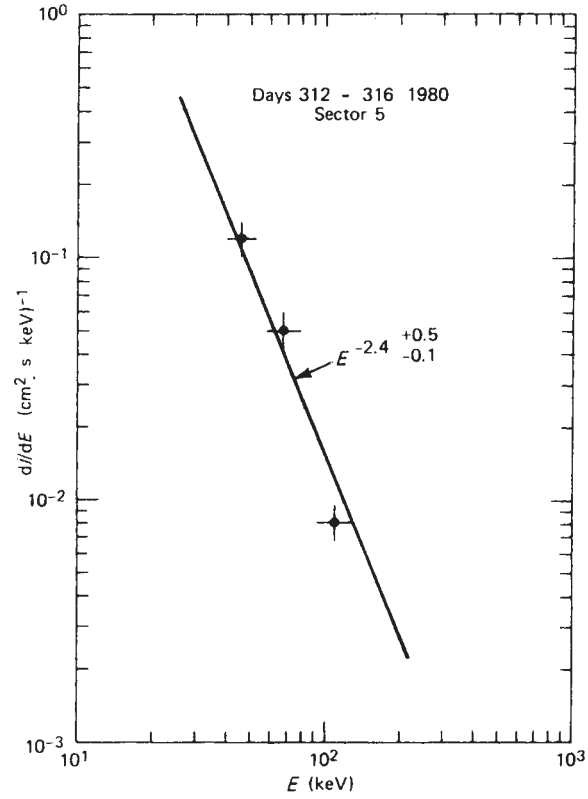


Fig. 4 Differential energy spectrum of the excess radiation, on the assumption that the detector response is due to energetic neutral hydrogen (Table 1).

softer spectrum than the flux of the neighboring sectors. As sectors 6, 7, 1, 2 and probably 3 are most likely to contain interplanetary particles, we compare sector 5 to only sector 4, in an attempt to estimate the additional flux in 5, as follows:

$$\Delta P_{1,2,3} = [(S_5 - S_4) - (Sq_5 - Sq_4)]_{1,2,3} \quad (1)$$

where $S_{5,4}$ are sector count rates and Sq_5, Sq_4 the quiet time sector rates averaged over days 298, 299 and 300. The excess count rates obtained by use of equation (1) are then normalized by multiplication with $(R/45\ R_S)^2$ and listed in Table 1. The statistical errors have been calculated as

$$(\bar{\sigma}_{1,2,3})^2 = \{(\overline{\delta S_4})^2 + (\overline{\delta S_5})^2 + (\overline{\delta Sq_4})^2 + (\overline{\delta Sq_5})^2\}^{1/2} \quad (2)$$

where

$$(\overline{\delta S_{4,5}})^2 = \frac{1}{5}\sum_{n=1}^{5} (\delta S_{4,5})_n^2 \left(\frac{R}{45 R_S}\right)_n^2 \quad (3)$$

with $n = 1$ to 5 representing days 312 to 316, 1980, and δSq_4, δSq_5 are the statistical errors of the quiet days average 298, 299 and 300. The errors in the count rate ratios in Fig. 3*b*,*c* shown for each day were calculated using an equation similar to equation (2). Table 1 contains the normalized count rates $\Delta P_{1,2,3}$, the corresponding fluxes and the statistical errors calculated with equation (2). Note that the r.m.s. scatter in the rates for ΔP_1 is larger than the statistical error, and is likely to be due to time variations in this softest part of the spectrum.

The averaged excess count rates in sector 5 shown in Table 1 are statistically significant. The additional counts could be due to planetary or ring X rays, energetic neutrals and/or interplanetary charged particles reflected off the bow shock, so that only upper limits for the neutral radiation can be derived. The averaged X-ray flux j_X can be calculated from the count rate by

$$j_X = \frac{\Delta P_{1,2,3}}{A \varepsilon_X \Delta E_X} \quad (4)$$

where A is detector area, ε_X is X-ray efficiency and ΔE_X is the electronic channel width. Similarly, the neutral particle flux j_n can be calculated from equation (4), but using the appropriate values for ε and ΔE. The results, normalized to a distance of 45 R_S from the planet, are included in Table 1.

Discussion

The significance of the fluxes listed in Table 1 must be examined in the context of processes known or expected to take place in the vicinity of Saturn. Production of X rays can occur in auroral processes recently observed to take place at Saturn[11] over a relatively narrow range in latitude, 78°–81.5°. If the electron precipitation takes place over the area of the observed UV signature ($A \sim 10^{19}$ cm^2, that is, 2% of the surface), then the emitted photon flux at the source is given by

$$j_X < \frac{j_X 4\pi R^2 \Delta E}{A} = 3.4 \times 10^6 \text{ photons cm}^{-2}\text{ s}^{-1} \quad (5)$$

with $\Delta E \sim 7$ keV, $R = 45\ R_S$. The corresponding electron flux

Table 1 Excess rates, fluxes in sector 5 (normalized to 45 R_S)

Day (normalization)*	ΔP_1(c.p.s.)	ΔP_2(c.p.s.)	ΔP_3(c.p.s.)
312 (8.1)	7.89×10^{-2}	1.66×10^{-1}	7.5×10^{-2}
313 (5.44)	7.38×10^{-2}	1.69×10^{-1}	
314 (3.4)	4.16×10^{-2}	1.38×10^{-1}	2.4×10^{-2}
315 (1.78)	2.80×10^{-2}	0.79×10^{-1}	2.1×10^{-2}
316 (1.0)	3.43×10^{-2}	0.76×10^{-1}	2.5×10^{-2}
Ave† (c.p.s.) ± 1σ	0.051 ± 0.008	0.126 ± 0.014	0.36 ± 0.012
Flux (counts cm^{-2} s^{-1})	0.64 ± 0.1	1.57 ± 0.18	0.46 ± 0.15
j_X(photons cm^{-2} s^{-1} keV^{-1})	5.4 ± 0.9	6.1 ± 0.7	
j_n(cm^2 s keV)$^{-1}$	0.12 ± 0.02	0.05 ± 0.01	0.008 ± 0.003
Energy range (keV)			
X rays	26–31	31–63	63–121
Neutrals/charged	40–53	53–85	85–139

* Factors in parentheses are obtained from $(R/45\ R_S)^2$ and are used to normalize the excess rate to 45 R_S.

† Averages normalized to 45 R_S.

© 1981 Nature Publishing Group

precipitating into the atmosphere can be estimated by using a conversion efficiency of $\sim 2\times10^{-5}$ photons per electron[12]; we find $J\sim 1.7\times10^{11}$ electrons $cm^{-2}\,s^{-1}$ which is some 10^3 larger than the maximum fluxes measured by the LECP experiment within the magnetosphere of the planet[13]. Hence, we conclude that the observed upper limit, if interpreted in terms of X rays, cannot be reasonably related to precipitating magnetospheric electrons. The same conclusion is reached even if precipitation is assumed to occur over the entire visible hemisphere.

Another possible X-ray source is the interaction of radiation belt electrons with the rings of Saturn, which extend out to $\sim 2.5\,R_S$. The surface area of the rings on the sunlit side is

$$S=(2.5\,R_S)^2\pi-(1\,R_S)^2\,\pi\simeq 6\times10^{20}\,cm^2 \qquad (6)$$

If we assume $\sim 1\%$ of the area to be filled with matter, the estimated X-ray production in the rings in the energy range 31–63 keV (PL02), is

$$J_X=\frac{6.1\times4\pi(43\,R_S)^2\times32}{6\times10^{18}}=2.7\times10^7\ \text{photons}\ cm^{-2}\,s^{-1} \qquad (7)$$

The corresponding electron precipitation is $\sim 1.4\times10^{12}\,cm^{-2}\,s^{-1}$, again $\sim 10^4$ larger than the maximum flux measured by the same instrument in the magnetosphere with this channel[13]. Note that there is a sharp cutoff in particle fluxes at the edge of ring A, as shown by Pioneer 11[14].

The Voyager 1 UV experiment[11] found that Saturn is surrounded by a torus of neutral hydrogen with a concentration $n_H\approx 10\,cm^{-3}$. The torus extends from $\sim 25\,R_S$ to $\sim 8\,R_S$ and its thickness could be $\sim 6\,R_S$. Thus charge exchange processes between radiation belt protons and the neutral hydrogen concentration can be expected. For a flux estimation we use the cross-sections given by Tinsley[5]. The product σv (cross-section × velocity of the particles) is as follows for the lowest three channels

PL01: (40–53 keV): $\sigma v_H^+\approx 3\times10^{-8}\,cm^3\,s^{-1}$

PL02: (53–85 keV): $\sigma v_H^+\approx 1.4\times10^{-8}\,cm^3\,s^{-1}$

PL03: (85–139 keV): $\sigma v_H^+\approx 3\times10^{-9}\,cm^3\,s^{-1}$

The lifetime against charge exchange is then

PL01: $\tau_1=(n_H\sigma v)^{-1}=(10\times3\times10^{-8})^{-1}=3.33\times10^6\,s$

PL02: $\tau_2=(n_H\sigma v)^{-1}=(10\times1.4\times10^{-8})^{-1}=7.14\times10^6\,s$

PL03: $\tau_3=(n_H\sigma v)^{-1}=(10\times3\times10^{-9})^{-1}=3.33\times10^7\,s$

The LECP instrument measured an average density outside the orbit of Rhea at $E\gtrsim 40$ keV of $\sim 5\times10^{-3}\,cm^{-3}$ (ref. 13). Taking the neutral hydrogen cloud volume of $\sim 2\times10^{33}\,cm^3$ (ref. 11) we can estimate the total energetic ion content as

$$N=nV=5\times10^{-3}\,cm^{-3}\times2\times10^{33}\,cm^3=10^{31}$$

The total energetic neutral particle production is given by

$$\dot N=\frac{nV}{\tau}=\frac{10^{31}}{10^7}=10^{24}\ \text{neutrals}\ s^{-1}$$

The loss rate may be obtained by estimating the total surface area of the neutral hydrogen disk,

$$A\simeq 2\pi[(25\,R_S)^2-(8\,R_S)^2]\simeq 1.3\times10^{23}\,cm^2$$

Thus energetic ($E\gtrsim 40$ keV) neutrals are lost at a rate

$$\frac{\dot N}{A}=\frac{10^{24}}{1.3\times10^{23}}\sim 8\,cm^{-2}\,s^{-1}$$

The flux at 45 R_S would be decreased by a factor $\sim(20/45)^2=0.2$, that is, $\dot N/A$ would be $\sim 1.6\,cm^{-2}\,s^{-1}$.

We now need to compare the expected flux at 45 R_S with the observations as summarized in Table 1. The differential flux from Table 1 has been plotted in Fig. 4; the form of the spectrum can be described by a power law of the form $E^{-\gamma}$ with $\gamma\sim 2.4$. To obtain the total flux we evaluate the integral

$$J=\int_{40\,keV}^{\infty}KE^{-2.4}\,dE$$

and find $J\sim 4.3\,cm^{-2}\,s^{-1}$. This value compares well with the expected value of $\sim 1.6\,cm^{-2}\,s^{-1}$ derived above.

We have demonstrated the existence of radiation above the detector background emanating from the vicinity of Saturn, upstream from the bow shock. The excess counts are most plausibly attributed to neutral radiation, that is, X rays and/or neutral energetic particles. Estimates of the number of energetic electrons necessary to produce the observed counts by precipitation in either the saturnian auroral regions or the rings are too high by factors $\sim 10^3$–10^4 over electron intensities observed within the magnetosphere of Saturn. Thus the observed excess counts are probably not due to X rays even if satellite surfaces are considered as possible X ray sources. We conclude that charge exchange of energetic ions with satellite tori is an important loss mechanism at Saturn as well as at Jupiter[15].

The efforts of many at APL/JHU, the Universities of Maryland and Kansas and Bell Laboratories contributed to this investigation. We thank our colleagues on the LECP team: T. P. Armstrong, W. I. Axford, C. O. Bostrom, E. P. Keath and L. J. Lanzerotti for various contributions, J. F. Carbary for assistance with graphics, and especially L. J. Lanzerotti for useful comments. This research was supported in part by NASA under Contract N00024-78-C-5384 between The Johns Hopkins University and the Department of the Navy, and under subcontract to the University of Maryland.

Received 13 May; accepted 7 July 1981.

1. Kremser, G. *Scientific Ballooning* (ed. Riedler, W.) 161 (Pergamon, Oxford, 1979).
2. Moritz, J., *J. geophys. Res.* **38**, 701 (1972).
3. Hovestadt, D., Haeusler, B. & Scholer, M. *Phys. Rev. Lett.* **28**, 1340 (1972).
4. Blake, J. B. *J. geophys. Res.* **81**, 6189 (1976).
5. Tinsley, B. A. *J. geophys. Res.* **81**, 6193 (1976).
6. Cheng, A. F. *Astrophys. J.* **242**, 812 (1980).
7. Kirsch, E., Krimigis, S. M., Kohl, J. W. & Keath, E. P. *Geophys. Res. Lett.* **8**, 169 (1981).
8. Krimigis, S. M. *et al. Space Sci. Rev.* **21**, 329 (1977).
9. Bridge, H. S. *et al. Science* **212**, 217 (1981).
10. Ness, N. F. *et al. Science* **212**, 211 (1981).
11. Broadfoot, A. L. *et al. Science* **212**, 206 (1981).
12. Evans, R. D. *The Atomic Nucleus* 619 (McGraw-Hill, New York, 1955).
13. Krimigis, S. M. *et al. Science* **212**, 225 (1981).
14. Van Allen, J. A., Thomsen, M. F., Randall, B. A., Rairden, R. L. & Grosskreutz, C. L. *Science* **207**, 415 (1980).
15. Cornwall, J. M. *J. geophys. Res.* **77**, 1756 (1972).

Topology of Saturn's main magnetic field

M. H. Acuña, J. E. P. Connerney & N. F. Ness

NASA/Goddard Space Flight Center, Laboratory for Extraterrestrial Physics, Greenbelt, Maryland 20771, USA

The Voyager 1 magnetic field observations at Saturn confirm the principally dipolar topology of the planetary magnetic field and suggest the need for more general models which incorporate non-potential field sources external to the planet and within the planetary magnetosphere.

THE close flyby of Saturn at 3.07 R_S (1 R_S = 60,330 km) by Voyager 1 in November 1980 provided a second opportunity to study *in situ* the intrinsic planetary magnetic field. The first detection of the Saturnian field was accomplished by experiments aboard Pioneer 11 during its close flyby of Saturn at 1.35 R_S in September 1979[1–4]. The most surprising

0028–0836/81/340721—04$01.00

© 1981 Macmillan Journals Ltd

characteristic derived from these magnetic field measurements was the near-alignment of the magnetic axis with the rotational axis of the planet. A small but non-zero axial displacement with respect to the planet's centre was also obtained. These features and the purely dipolar nature of the planetary field have significant implications for the general problem of the generation of magnetic fields by massive rotating objects.

The detection of Saturn at kilometric and hectometric wavelengths by the Voyager planetary radio astronomy experiment and identification of periodicities at 10:66 indicate that emissions originate from a restricted range of planet-centred longitudes[5]. This constitutes a paradox in terms of what would be intuitively expected from an axisymmetrical magnetic field, assuming that the emissions originate near the feet of the field lines. A more detailed analysis of the data obtained by Voyager 1 during its encounter with Saturn is presented here, with the emphasis on the uncertainties associated with the determination of model parameters describing the main magnetic field. The results of a preliminary analysis[6] use data available immediately after the time of the encounter. Since then determinations of spacecraft attitude, position and alignment of the magnetometer sensors with respect to the main body of the spacecraft have been made, which can have a direct impact on the models derived from the observations. The independent results obtained from Voyager 1 allow a more general interpretation of Pioneer 11 observations and provide an insight into the phenomena taking place in the saturnian magnetosphere.

The flyby trajectories for Pioneer 11 and Voyager 1 were such that different ranges of planet-centred latitudes, longitudes and corresponding radial distances were sampled during each encounter. These characteristics are illustrated in Fig. 1 where the sub-spacecraft latitude and longitude for each trajectory during the closest approach phase are shown. The planetocentric radial distance to each spacecraft is indicated as well as the spacecraft event time (SCET) for Voyager 1. Figure 1 shows that the Pioneer 11 trajectory is largely equatorial, remaining within ±6° of the planetographic equator for $r < 8\ R_S$.

In contrast, Voyager 1 approached Saturn from the south, reaching a maximum latitude of −41° at a radial distance of $3.14\ R_S$ just before periapsis at $3.07\ R_S$. Voyager 1 crossed Saturn's equatorial plane into the northern hemisphere at $6.3\ R_S$ while outbound from periapsis. The larger latitude coverage of Voyager 1 is important to the modelling process, primarily because it allows a more accurate determination of hemispherical asymmetries, such as the polar offset derived from the Pioneer measurements[2,4]. Note also in Fig. 1 the narrow range of planet centred longitudes sampled by Pioneer 11 inside $6\ R_S$ (~120°). Voyager 1, in addition to the greater latitude coverage, extends this range to ~135°, although it remains entirely in the southern hemisphere within the same radial distance range.

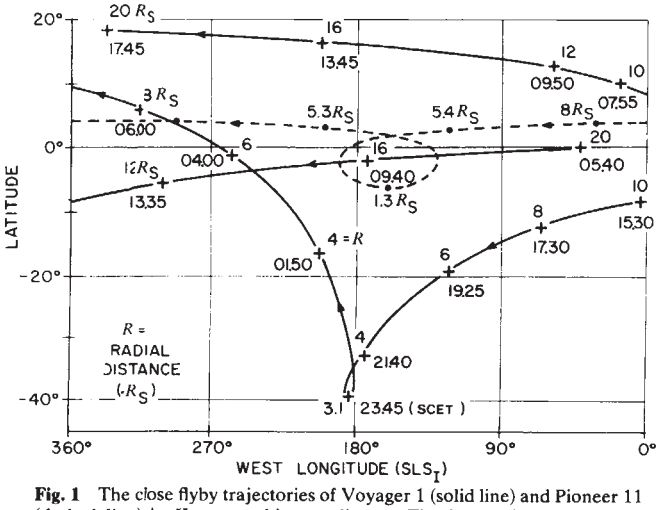


Fig. 1 The close flyby trajectories of Voyager 1 (solid line) and Pioneer 11 (dashed line) in Kronographic coordinates. The locus of sub-spacecraft points in SLS-longitude and latitude is shown. Voyager 1 crossed the equatorial plane from north to south at 20 R_S, reaching a maximum latitude of −40°. In contrast the Pioneer 11 trajectory remained within ±6° of the equator. SCET, spacecraft event time.

The analysis of magnetic field observations along a spacecraft flyby trajectory has been described elsewhere and presents several problems related to the uniqueness and uncertainties of parameters in the mathematical models derived from the measurements[7-9]. The usual assumption made is that the observations are performed in a region that is current-free, that is, external (or internal) to the sources of the measured magnetic field ($\nabla \times \vec{B} = \mu_0 \vec{J} = 0$). In these conditions, the field can be represented as the gradient of a scalar potential function V, as $\vec{B} = -\nabla V$. The scalar potential has traditionally been represented by means of spherical harmonic functions with suitably normalized coefficients[10]. Two potential functions are normally used, one representing sources which reside in the interior of the planet and a second one which accounts for sources outside the region where the measurements are obtained. This approach was used in the preliminary report on the Voyager 1 magnetic field observations at Saturn[6]. The condition of $\nabla \times \vec{B} = \mu_0 \vec{J} = 0$ is reasonably well satisfied within the radial range $R < 6R_S$. 48-s averages of magnetic field data were used to derive a least-squares model fit to the observations using an interior, centred, tilted dipole and an exterior uniform field (I1E1 spherical harmonic model with internal and external components both first order). These results yielded the Schmidt normalized coefficients reported by Ness *et al.*[6], that is: $g_1^0 = 0.209$ G, $g_1^1 = 0.00220$ G, $h_1^1 = 0.00124$ G; external $G_1^0 = -8.8 \times 10^{-5}$ G, $G_1^1 = -4.2 \times 10^{-5}$ G and $H_1^1 = 3.3 \times 10^{-5}$ G. These results agree well with those reported by the Pioneer 11 investigators[1-4]. With respect to the dipole displacement from the centre of the planet, the preliminary Voyager 1 studies placed an upper limit of $0.02\ R_S$ to any possible offset. The most significant new results were that the $0.04\ R_S$ northward offset derived from the Pioneer 11 measurements is not reproduced in the Voyager 1 models, and that a definite dipole tilt of ~0.7° at a SLS[11] longitude of 331° was obtained. Note, however, the assumptions made in the model: current-free observations as well as a uniform external field.

Additional studies of the magnetic field data obtained during the traverse of the middle magnetosphere of Saturn ($6\ R_S \leqslant R \leqslant 16\ R_S$) have led Connerney *et al.*[12] to conclude that Saturn is surrounded by an eastward-flowing ring current which extends from 8.5 to $15.5\ R_S$ in the equatorial plane and with a total thickness of $\sim 5R_S$. The current density in this sheet decreases proportionally to distance from the planet and is an order of magnitude smaller than that found at Jupiter. A more realistic magnetic field model for the analysis of Saturn's main field in which the distributed ring currents are explicitly modelled is constructed as follows: an I1E1 model is first least-squares fitted to the observations obtained within $\sim 6\ R_S$ of close approach. These internal field parameters constitute a first approximation to Saturn's main field: the external terms (uniform field) approximate the field due to magnetospheric currents. The internal field is then subtracted from observations obtained within $\sim 30\ R_S$ and the resulting external field is modelled in terms of an azimuthal ring current. The (solenoidal) field of the ring currents is then adopted as a more accurate external field representation, and assumed in a least-squares fit to determine new internal field coefficients (I1E0).

This procedure does not necessarily converge to a unique set of model parameters and each iteration step must be guided by appropriate physical constraints as well as results from correlative investigations. The results of such a study are given in Table 1. As done previously, 48 s-averaged field values obtained for $R < 8R_S$ have been used. Two weighting schemes have been used, one in which uniform weights are attached to the measurements ($\omega_i \equiv 1$) and a second one in which the weights have been chosen inversely proportional to the magnitude of the measured field ($\omega_i \propto [0.01 \times |\vec{B}_i|]^{-1}$). The first is appropriate for observations assuming a uniform (random) noise level and the second for observations assuming a noise component which is

© 1981 Nature Publishing Group

Table 1 Saturn's magnetic field models from Voyager 1 observations

Parameter	Case 1 I1E1*	2 I1E1	3 I1 + Ring curr.†	4 I1E1	5 I1 + Ring curr.†
g_1^0	20,900	21,121	20,926	20,940	20,966
g_1^1 (nT)	220	324	401	230	276
h_1^1	124	43	152	170	191
G_1^0	−8.8	−11	—	−9	—
G_1^1 (nT)	−4.2	−1	—	−3	—
H_1^1	3.3	−1	—	1	—
r.m.s. residual (nT)	2.7	2.0	1.9	2.7	2.8
Dipole tilt, θ_D	0.7°	0.89°	1.17°	0.78°	0.92°
Longitude, λ_D [SLS]	331°	352°	339°	324°	325°
Radial distance range of data	$<6R_S$	$<8R_S$	$<8R_S$	$<8R_S$	$<8R_S$
Weights	1	$[.01\times B_i]^{-1}$	$[.01\times B_i]^{-1}$	1	1

* Ref. 6.
† Ring current model of Connerney *et al.*[12]

proportional to the local field magnitude[9]. The derived tilt angles and longitude of the tilt are also indicated in Table 1. The r.m.s. of the residuals of the model fits show that all these models fit the observations very well. As discussed by Ness *et al.*[6], second-order (quadrupole) models do not provide a significant reduction in the residuals. The implication of this is that the derived offset of the dipole from the centre of the planet is small and consistent with zero to the level that can be ascertained from the magnetic field measurements alone. Note from Table 1 that models incorporating the field due to the ring current explicitly, rather than the I1E1 models which represent this field as uniform, do not provide a significant reduction in the r.m.s. of the residuals. This is due to the model residuals being dominated by a localized and presumably non-potential feature unrelated to the ring current or internal field.

We now illustrate the residuals in the three orthogonal field components, ΔB_R, ΔB_θ, ΔB_ϕ after the internal field model plus that due to the modelled ring current are subtracted from the observations. This is shown in Fig. 2 for case 5 in Table 1. This particular model was chosen primarily for convenience. In general, the magnitude of the residuals is small, <10 nT. Some still unresolved engineering and attitude determination problems are evident in Fig. 2. The 'disturbance' at 00:43 UT and extending until ~01:00 is associated with a spacecraft roll manoeuvre and the lack of sufficiently accurate knowledge of spacecraft and therefore magnetometer sensor orientation during the manoeuvre. The sensor alignment with respect to the spacecraft reference axis was determined by means of an onboard calibration coil[13] and the analysis of spacecraft manoeuvres in steady-state fields during both cruise and planetary encounters at Jupiter and Saturn. This orientation is known to better than ±0.2°. Also evident in Fig. 2 are small discontinuities associated with instrument range changes, the updating and interpolation of spacecraft attitude matrices and analogue-to-digital converter digitization uncertainties. In addition, the spacecraft limit cycle (±0.06°), which is unmodelled, contributes to the observed noise where the ambient field is large (1 nT in 10^3 nT).

The largest contribution to the residuals occurs between 21:00 on day 317 and 05:00 on day 318, particularly in the radial component, ΔB_R. An additional feature can be seen superimposed on ΔB_R and ΔB_θ curves occurring between 22:30 on day 317 and 01:00 on day 318. This quasi-abrupt enhancement of the residuals seems to be associated with a magnetospheric current system quite distinct from the azimuthal current system already modelled. This feature was observed as Voyager 1 passed through the dusk sector of Saturn's magnetosphere, nearly coincident with closest approach. In planetocentric coordinates, the spacecraft was moving primarily in latitude at this time (Fig. 1), actually reversing momentarily its westward motion, as viewed from Saturn.

The residuals of internal magnetic field models are dominated by a highly correlated, very localized and presumably non-potential field contribution near periapsis. Thus the current representations of the field are not sufficient to describe the observed field to better than 10 nT. This includes models explicitly incorporating distributed azimuthal currents as well as potential field models. Note that the residuals illustrated in Fig. 2 are evidence of local currents but the plotted field components are not the magnetic field signature of such currents. The plotted residuals are only that part of the observed field that cannot be expressed with the limited models available. Certainly some part of the field contributed by local currents has been absorbed by the internal field coefficients and therefore does not appear in the residual plot. The problem of separating the different contributions to the observed field due to the planetary field, the ring current, and the other local currents is the primary challenge to be met in obtaining better estimates of Saturn's internal field. To do so requires a model of the inner magnetosphere in which the very localized currents are explicitly taken into account.

Several model current systems that might be expected on physical grounds to be active in this region of Saturn's magnetosphere are being considered. For example: (1) field-aligned currents associated with the interaction of Saturn's corotating magnetosphere with a satellite, or particle ring, reminiscent of the currents linking Jupiter and its satellite Io, as observed by Voyager 1[14]. The most promising candidates here seem to be the E-ring and the satellite Dione. Voyager passed ~20,000 km north of the field line linking Dione at ~00:10, day 318, as it traversed the dusk sector. (2) A dawn–dusk system of Birkeland currents driven by the solar wind-induced asymmetry. (3) A system of Birkeland currents driven by a non-azimuthally symmetric torus or an as yet unidentified surface magnetic anomaly similar to that proposed for Jupiter[15]. An enhancement of high energy (1–2 keV) electron fluxes observed by the Voyager 1 plasma experiment in the dusk sector,

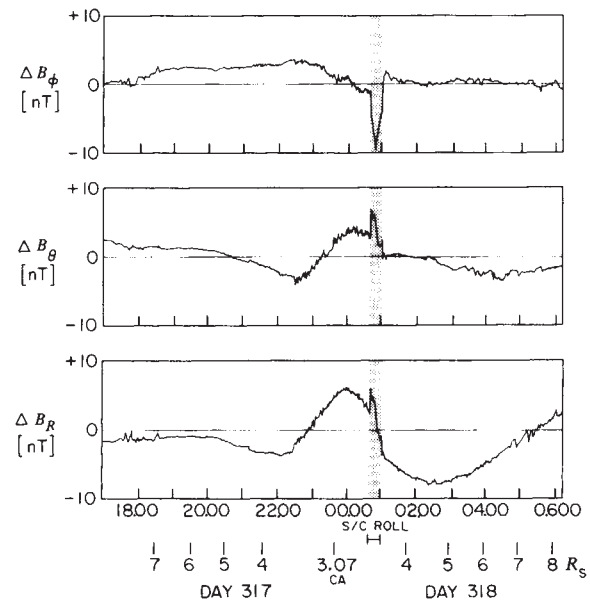


Fig. 2 Plot of the component residuals, ΔB_R, ΔB_θ and ΔB_ϕ obtained after subtracting a dipole model plus the field from the ring current as determined from the Voyager 1 observations. The abrupt enhancement observed between 22.30 and 01.00 h of day 318 is interpreted as evidence for a localized current system flowing in the inner saturnian magnetosphere.

© 1981 Nature Publishing Group

concomitant with the magnetic field signature, has been described by Sittler *et al*[16]. As additional correlative observations become available, we hope to offer a more detailed model.

Of the models in Table 1, cases 2 and 3, derived using weights inversely proportional to the local field magnitude, minimize the effect of the localized current system encountered near periapsis. This unmodelled 10-nT perturbation is less likely to affect significantly the internal field coefficients in the weighted least-squares fits compared with the unweighted cases, as larger errors are tolerated in higher fields. Case 3 in Table 1 is regarded as best representing the global characteristics of the saturnian field, as deduced from Voyager 1 observations alone.

To estimate model parameter uncertainties we need to assess the impact on the internal field coefficients of the unmodelled current system. Lacking a detailed model, we can estimate the parameter uncertainties by comparing models relatively unaffected by the currents (weighted fits) with those that are more strongly dependent on the residuals near closest approach (unweighted fits). Such a comparison suggests a dipole moment of 0.21 ± 0.005 G$-R_S^3$, tilted $1.0 \pm 0.3°$ towards an SLS longitude of $340° \pm 20°$. The northward displacement of the dipole by $0.04\,R_S$ deduced from the Pioneer 11 observations is not substantiated by the Voyager 1 observations, although we must take into account that the near-equatorial trajectory of Pioneer 11 was poorly suited for detecting hemispherical asymmetries. Voyager 2 will fly by Saturn in August 1981; its close encounter trajectory will provide a latitude coverage of $\sim \pm 30°$ within $8\,R_S$ and facilitate a more accurate estimate of a possible dipole polar offset.

Certain aspects of Saturn's main magnetic field relevant to the planet's interior have been discussed by Stevenson[17]. In particular, the unexpectedly small dipole moment seems to be consistent with the gravitational settling of helium, which leads to a much smaller electrically conducting and convecting region than would be expected of a homogeneous distribution of hydrogen and helium. The IR studies by Voyager 1 have demonstrated such an atmospheric depletion of helium, obtaining an 11% helium mass fraction at Saturn as compared with the 19% observed at Jupiter and expected on a cosmochemical basis[18]. The near alignment of Saturn's magnetic and rotational axes is likewise consistent with the differential rotation of the conducting core in Stevenson's model of Saturn's interior[17].

We thank our colleagues on the Voyager team, especially J. Scudder and E. Sittler of the plasma experiment for helpful discussions. We also thank R. Thompson and F. Ottens for the analysis of these data.

Received 13 May; accepted 10 June 1981.

1. Acuña, M. H. *et al. Science* **207**, 444–446 (1980).
2. Smith, E. J. *et al. Science* **207**, 407–410 (1980).
3. Acuña, M. H. *et al. J. geophys. Res.* **85**, 5675–5678 (1980).
4. Smith, E. J. *et al. J. geophys. Res.* **85**, 5655–5674 (1980).
5. Kaiser, M. L. *et al. Science* **209**, 1238–1240 (1980).
6. Ness, N. F. *et al. Science* **212**, 211–217 (1981).
7. Acuña, M. H. & Ness, N. F. in *Magnetospheric Particles and Fields* (ed. McCormack, B. M.) 311–323 (Reidel, Dordrecht, 1976).
8. Davis, I., Jr & Smith, E. J. in *Magnetospheric Particles and Fields* (ed. McCormack, B. M.) 301–310 (Reidel, Dordrecht, 1976).
9. Connerney, J. E. P. *J. geophys. Res.* (in the press).
10. Chapman, S. J. & Bartels, J. *Geomagnetism* (Oxford University Press, 1940).
11. Desch, M. D. & Kaiser, M. L. *Geophys. Res. Lett.* **8**, 253–256 (1981).
12. Connerney, J. E. P. *et al. Nature* **292**, 724–726 (1981).
13. Behannon, K. W. *et al. Space Sci. Rev.* **21**, 235–257 (1977).
14. Acuña, M. H. *et al. J. geophys. Res.* (in the press).
15. Dessler, A. J. & Vasyliunas, V. M. *Geophys. Res. Lett.* **6**, 37–40 (1979).
16. Sittler, J. C., Scudder, J. C. & Bridge, H. S. *Nature* **292**, 711–714 (1981).
17. Stevenson, D. J. *Science* **208**, 746–747 (1980).
18. Hanel, R. *et al. Science* **212**, 192–200 (1981).

Saturn's ring current and inner magnetosphere

J. E. P. Connerney, M. H. Acuña & N. F. Ness

NASA/Goddard Space Flight Center, Laboratory for Extraterrestrial Physics, Greenbelt, Maryland 20771, USA

The Voyager 1 magnetic field observations at Saturn reveal an equatorial system of (eastward) azimuthal currents, very similar in certain respects to that responsible for the jovian magnetodisk.

IN November 1980, Voyager 1 became the second spacecraft to obtain *in situ* observations of the magnetosphere of Saturn: little more than a year earlier, the Pioneer 11 spacecraft returned the first evidence of Saturn's magnetosphere. Analyses of the Pioneer 11 magnetic field observations[1,2] revealed a dipolar planetary magnetic field of moment 0.21 G$-R_S^3$, rather less than that expected on the basis of scaling laws[3,4]. [The units G$-R_S^3$ are dimensionally equivalent to standard units of magnetic dipole moment, the numerical value having been divided by the cube of the planetary radius.] More surprising was the near alignment of Saturn's rotational and magnetic dipole axes, a unique characteristic among the known planetary dynamos and therefore of central interest in dynamo theory[5]. A preliminary estimate of 0.7° angular separation of Saturn's rotational and magnetic dipole axes has recently been derived from the Voyager 1 magnetic field observations[6].

Saturn's lack of an appreciable dipole tilt results in an observational limitation insofar as studies of magnetospheric phenomena *in situ* by spacecraft are involved. Figure 1 shows the trajectories of each spacecraft to encounter Saturn (Pioneer 11 (P11) in September 1979; Voyager 1 (V1) in November 1980; and Voyager 2 (V2) in August 1981) in a cylindrical (ρ, z) planet centred coordinate system, which in this case is practically identical to the magnetic equatorial coordinate system. Each spacecraft remains essentially at constant magnetic latitude for a large percentage of the encounter, with only one (V2) or two (V1, P11) crossings of the magnetic equator. Particularly evident is the rather singular nature of Pioneer 11's trajectory, which traced a nearly identical path outbound from periapsis as inbound. In contrast to the periodic variations in the magnetic

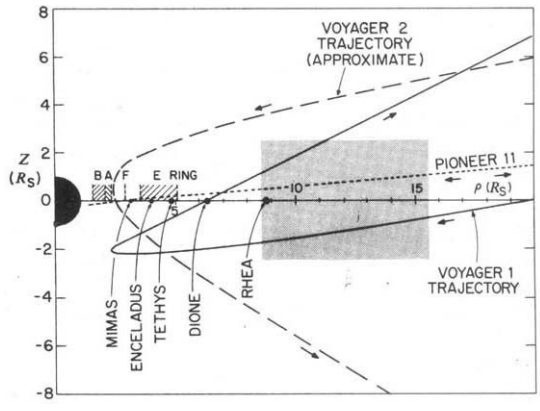


Fig. 1 Trajectories of Pioneer 11, Voyager 1 and Voyager 2 at Saturn in a cylindrical planetocentric equatorial coordinate system. Positions of the major satellites and rings are indicated; satellite diameters are exaggerated by a factor of 10. Stippled region is the region of (model) distributed currents in Saturn's magnetosphere.

0028–0836/81/340724—03$01.00

© 1981 Macmillan Journals Ltd

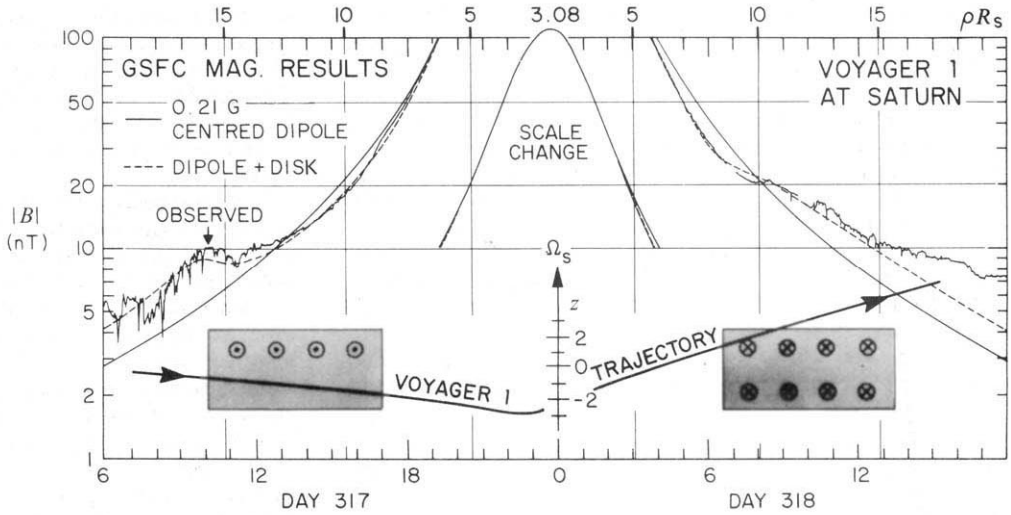


Fig. 2 Magnitude of observed magnetic field within ~20 R_S of Saturn and comparison with an internal dipole model (solid line) and a model containing a dipole and distributed ring currents (dashed line) discussed in test. The Voyager 1 trajectory is illustrated beneath these observations, showing distance from the equator as a function of time. Note the different scale used for vertical dimension of trajectory insert.

field and charged particle environment observed in Jupiter's magnetosphere concomitant with the regular variation of magnetic latitude due to that planet's rotation and dipole tilt (9.6°), at Saturn, the observations reflect more the systematic co-variation of ρ and z.

The Voyager 1 magnetic field observations at Saturn are shown in Fig. 2, in which the magnitude of the observed magnetic field is compared with that of a 0.21 G$-R_S^3$ centred dipole. The departure from a dipole is noticeable during the approach to periapsis at radial distances exceeding ~8 R_S, whereas outbound the differences are noticeable at radial distances in excess of ~5R_S. The observed field magnitude is appreciably less than that of the model dipole at small radial distances ($\rho < \sim 10R_S$) and greater than the model dipole in the more distant magnetosphere. These characteristics can be understood by introducing a model current system similar to that originally applied to observations of the jovian magnetodisk[7]. In this model, large-scale azimuthal currents flow eastward in an annular disk extending from 8.5 R_S to 15.5 R_S in Saturn's equatorial plane. The current is distributed uniformly in $\hat{z}$ over the total disk thickness of 5R_S and decreases with radial distance from Saturn as $1/\rho$. The physical dimensions of this annulus roughly correspond to those of the 'extended plasma sheet' observed by the Voyager 1 plasma experiment[8]. Because the radial extent of the disk, 7 R_S, is almost equal to the thickness, and considering the radial decrease in model current density, we refer to this phenomenon as a ring current rather than a magnetodisk to distinguish it from the more two-dimensional geometry of the jovian azimuthal currents. The model illustrated in Fig. 2 has a current density of $\sim 0.3\times10^6$ A/R_S^2 (~0.1 mA km^{-2}) at the inner edge of the annulus, decreasing to $\sim 0.15\times10^6$ A/R_S^2 at its outer edge. These current densities are about an order of magnitude less than those observed at Jupiter.

The derived model is a result of a vector fit to the Voyager observations and self-consistently satisfies both of the observed radial and vertical field components. The magnetic field of this equatorial system of currents is largely solenoidal. The near axis field is essentially vertical and oppositely directed to Saturn's (equatorial) main field; at larger distances the two add. Above the plane of symmetry of the distributed current system, the radial field is positive outward and below the equator, positive inward, increasing approximately linearly with distance, Z, from the equatorial plane for $Z < D$. Thus the field lines are radially stretched in a manner reminiscent of the stretching or drawing out of field lines in the near equatorial regions of the jovian magnetodisk.

The total azimuthal current, and the distribution of the current in $\hat{\rho}$, is inferred from the radial dependence of the external vertical field component. The external field is estimated by subtracting from the observations a best-fitting dipolar (0.21 G$-R_S^3$) internal field. The distribution of that current in $\hat{z}$ is inferred from the variation of the radial component of the external field. While the Voyager trajectory at Saturn did not permit numerous traversals of the current-carrying region as at Jupiter, it did provide observations within the distributed currents sufficient to allow the localization of source currents in Saturn's magnetosphere. An earlier model[6], characterized by currents confined to an annular disk of radial extent 8–16 R_S, with a total thickness of 4R_S, fits the observations about as well if the current density is increased such that the linear current density in Saturn's equatorial plane remains unchanged.

The near-axis field of the azimuthal currents is ~10 nT, which agrees well with the external field inferred from spherical harmonic analysis[6,9] of the Voyager 1 encounter observations. The field of this model current system reaches a maximum (12.5 nT) along the trajectory as Voyager 1 approaches the inner edge of the annulus at hour 6, day 318 while outbound from periapsis. The field enhancement observed as Voyager 1 approached the outer edge of the annulus at 15.5 R_S while inbound to periapsis is understood as an edge effect. The lack of such a localized enhancement of the modelled field during the outbound traversal is a simple consequence of Voyager 1's higher latitude at this time. Beyond 15 R_S outbound the discrepancy between the model field and the observations grows as the contribution of unmodelled cross-tail currents[10] to the total field grows.

A model of Saturn's magnetosphere, constructed by superposition of the dipolar 0.21 G$-R_S^3$ internal field and that of the model ring currents, is compared in Fig. 3 with a dipolar magnetic field. The radial stretching of field lines in the equatorial plane and the non-dipolar nature of the model field is evident at radial distances exceeding 5 R_S. The global character of the field due to the azimuthal currents ensures a departure from a dipolar magnetic field configuration at high latitudes as well. This axisymmetric model magnetosphere neglects the

Table 1 Times at which Voyager 1 encountered field lines which pass through the equator at a radial distance occupied by a known absorber

Feature	Radial distance	(Nearest minute) Model MS			Dipole	
Inbound		Day	h	min	h	min
Rhea	8.75	317	17	26	17	7
Dione	6.28	317	20	32	20	17
E Ring Outer Edge	5.0	318	0	7	0	2
Tethys	4.92	318	0	11	0	7
Outbound						
Tethys	4.92	318	2	45	2	46
E Ring Outer Edge	5.0	318	2	52	2	53
Dione	6.28	318	4	18	4	18
Rhea	8.75	318	6	26	6	35

©1981Nature Publishing Group

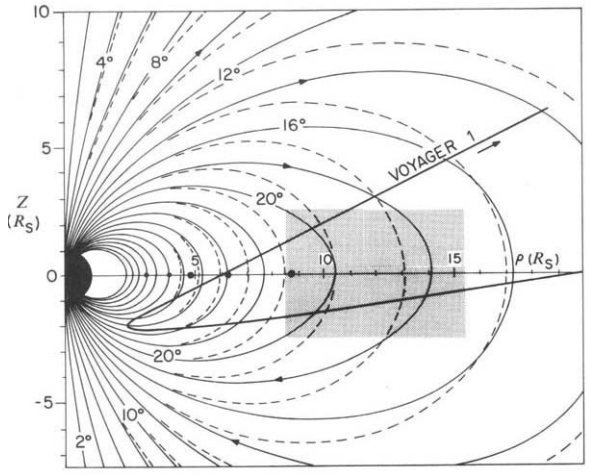


Fig. 3 Meridian plane projection of magnetosphere field lines for a dipolar field model (dashed) and a model containing a dipole and distributed ring currents (solid) discussed in the text. Field lines are drawn for 2° increments in co-latitude. Positions of the major satellites Mimas, Enceladus, Tethys, Dione, and Rhea are as in Fig. 1.

contribution of magnetopause and tail current systems driven by the solar wind interaction which become non-negligible in the more distant magnetosphere ($R > 15\ R_S$).

Saturn's ring current has important implications for charged-particle motion in Saturn's magnetosphere, particularly the absorption of trapped radiation by its many satellites and rings. Satellite absorption is conceptually simple: all charged particles that, in the course of their mirroring or latitudinal bounce motion, intersect an absorber are removed from the local particle population. In practice, charged particle absorption signatures associated with satellite sweeping are energy and species dependent, reflecting particle drift motion and finite gyroradius effects as well as particle diffusion. Thus, a complete description of charged particle motion relevant to satellite absorption requires a knowledge of the magnetic field line geometry and field gradient and curvature drifts over the particle's bounce path.

The absorption signature observed by the Voyager 1 cosmic ray experiment[11] near the orbital position of Rhea illustrates well the effects of Saturn's ring current on charged particle trajectories. Indeed, the premature occurrence of the absorption signature observed by Voyager 1 outbound in the vicinity of Rhea was cited as evidence of non-dipolar deformations of Saturn's post-midnight magnetosphere[11]. Given the model magnetosphere deduced from the magnetic field observations, we have traced field lines from the Voyager trajectory to the equatorial plane to arrive at the results presented in Table 1, times at which Voyager 1 encountered field lines which pass through the equator at a radial distance occupied by a known absorber. Also listed are the corresponding times assuming a dipolar model magnetosphere. The actual time at which the Rhea absorption signature was observed was ~06:25, day 318, which agrees very well with the model results presented here and ~10 min before that expected in a dipolar field.

The Rhea absorption signature was observed while Voyager 1 passed ~4° east of Rhea's orbital position. Magnetospheric plasma corotates from Rhea to Voyager 1 in ~7 min; in the corotating reference frame protons drift eastward, in the corotational direction, and electrons drift westward. Observation of an absorption signature in the 0.6-MeV electron data[11] was interpreted as evidence that the detector was responding to lower energy electrons, as electrons of energy >0.57 MeV (60° pitch angles) have drifted sufficiently in a dipole field that they trail Rhea in its orbit about Saturn. While we have not performed the required bounce-averaged gradient and curvature drift computation using our model field, we note that in the vicinity of Rhea's orbit Saturn's magnetosphere is sufficiently non-dipolar so as to cast doubt on longitudinal drifts computed with a dipolar field. The equatorial field gradient at Rhea's orbital distance is only ~0.6 that of a dipole field, and the local field magnitude is also ~0.66 that of a $0.21\ \mathrm{G} - R_S^3$ dipole. Particles of a particular energy with 90° pitch angles will thus drift with ~50% greater velocity than identical particles in a dipole field. In addition, the variation of the longitudinal drift rate with pitch angle is greater in the ring current model field than in a purely dipole field. Finally, we note that pitch angles derived assuming a dipole field geometry can be substantially in error (~20°) in the region of the magnetosphere where the Rhea absorption feature was observed. We conclude that in the cosmic ray experiments the 0.6 MeV detector may well have been responding to electrons with energies >0.6 MeV, as designed, and not to electrons below the nominal detection threshold of that instrument.

The total integrated ring current in Saturn's magnetosphere is $\sim 3 \times 10^6$ A, only a few per cent of the integrated azimuthal current flowing in the jovian magnetodisk. The near axis field of Saturn's ring current is ~10 nT; the ratio of this quantity, $B_r(0)$, to Saturn's equatorial dipole field, B_e, is $\approx 4.8 \times 10^{-4}$. The same quantity computed from a model of Jupiter's azimuthal disk currents[7] using $B_r(0) \sim 200$ nT and $B_e = 4.2$ G, is also $\sim 4.8 \times 10^{-4}$. This implies[12,13] that the total kinetic energy of trapped particles in the magnetosphere (E) scales with the available dipole magnetic energy outside the planet's surface (U). We note that while the ratio E/U is the same for both of the giant planets, Saturn's total E is little more than 0.15% of Jupiter's. The same ratio, applied to the Earth's magnetosphere (assuming $B_e \sim 0.3$ G), leads to a near-axis ring current field of ~14.5 nT. This is well within the 10–23 nT range estimated for the near-axis field of the Earth's quiet ring current on the basis of models fitted to Imp magnetic field observations[14]. That three so vastly different magnetospheres apparently have almost the same E/U ratio is indeed surprising and suggests a limiting process whereby such a ratio is maintained.

Saturn's magnetosphere is well approximated by an axisymmetric, but non-dipolar field within ~15 R_S of the planet. The non-dipolar nature of Saturn's magnetosphere is evident at radial distances exceeding ~5 R_S and can be expected to have a significant role in charged-particle motion in Saturn's magnetosphere. Beyond 15 R_S, the asymmetries introduced by the solar wind interaction become non-negligible, particularly the formation of a magnetotail in the night-side magnetosphere. We suspect that Saturn's ring current is a permanent feature of Saturn's magnetosphere, much like the jovian magnetodisk. A better understanding of possible temporal characteristics of Saturn's magnetosphere must await analysis of Pioneer 11 observations already obtained and those to be obtained by Voyager 2 in August 1981.

We thank our Voyager colleagues and colleagues at Goddard Space Flight Center, particularly T. J. Birmingham and D. P. Stern, for useful discussions.

Received 13 May; accepted 10 June 1981.

1. Smith, E. J. *et al. J. geophys. Res.* **85**, 5655–5674 (1980).
2. Acuña, M. H., Ness, N. F. & Connerney, J. E. P. *J. geophys. Res.* **85**, 5675–5678 (1980).
3. Busse, F. *Solar System Plasma Physics* Vol. 2 (ed. Kennel, C. F., Lanzerotti, L. J. & Parker, E. N.) 293–317 (North-Holland, Amsterdam, 1979).
4. Dolginov, Sh. Sh. *Rep. Izmiran*, Moscow (1978).
5. Todoeschuck, J. P. & Rochester, M. G. *Nature* **284**, 250–251 (1980).
6. Ness, N. F. *et al. Science* **212**, 211–217 (1981).
7. Connerney, J. E. P., Acuña, M. H. & Ness, N. F. *J. geophys Res.* **86**, (in the press).
8. Bridge, H. S. *et al. Science* **212**, 217–224 (1981).
9. Acuña, M. H., Connerney, J. E. P. & Ness, N. F. *Nature* **292**, 721–724 (1981).
10. Behannon, K. W., Connerney, J. E. P. & Ness, N. F. *Nature* **292**, 753–756 (1981).
11. Vogt, R. E. *et al. Science* **212**, 231–234 (1981).
12. Dessler, A. J. & Parker, E. N. *J. geophys. Res.* **64**, 2239–2252 (1959).
13. Carovillano, R. L. & Siscoe, G. L. *Rev. Geophys. Space Phys.* **11**, 289–353 (1973).
14. Mead, G. D. & Fairfield, D. H. *J. geophys. Res.* **80**, 523–534 (1975).

© 1981 Nature Publishing Group

Arc structure in Saturn's radio dynamic spectra

A. Boischot*, Y. Leblanc*, A. Lecacheux*, B. M. Pedersen* & M. L. Kaiser†

*Observatoire de Paris, F-92190 Meudon, France
†Goddard Space Flight Center, Code 695, Greenbelt, Maryland 20771, USA

The dynamic spectra of Saturn's kilometric radiation show arc structures whose main characteristics are described and compared with those observed in the decametre emission of Jupiter. The origin of the arcs is probably similar for the two planets.

ONE of the most spectacular findings of the Voyager 1 planetary radio astronomy experiment (PRA) was the discovery of nested arc-like structures in the dynamic spectra of Jupiter's decametric emission. These arcs have curvature in either a direction towards increasing time (vertex early arcs or VEA) or decreasing time (vertex late arcs or VLA)[1,2]. Several explanations for their structure have been proposed[3-7]. Similar arc structures are also evident in Saturn kilometric radiation (SKR) (see Fig. 1). These structures appear superimposed on the strong intensity modulation which is controlled by the rotation of the planet (sub-Voyager longitude in SLS) and on fast and narrow band fluctuations which give a very large variability to the PRA observations on a 6-s, 20-kHz scale. This very fine structure seems to be always present.

Characteristics of the arcs

The arc structure for Saturn, although easy to recognize, is not as regular as it is for Jupiter. This is due to the superposition of fine structure, and also because the amplitude of modulation of the emission by the arcs is small, only a few decibels on average, compared with 10 or 20 db for Jupiter. Thus for each rotation only a few arcs are clearly distinguished. The arc structure extends over the whole frequency range of the emission, but is often more clearly seen near the high frequency limit. Some phenomenon (possibly absorption or refraction) probably fixes the low frequency limit of the observed emission, and this prevents the lower legs of the arcs being seen to their full extent.

During the month around encounter the shapes of the arcs do not change much with time or SLS (Fig. 2). We can draw an 'average arc' with its vertex around 400 kHz and symmetrical curvature (on a linear frequency scale) around this vertex. The total extent is about 20 min and the shape is the same for VEA and VLA. In several cases, successive arcs are separated by 10–30 min, so that they appear to be nested like jovian arcs, but this nesting is not as regular as for Jupiter.

The arc structure could be seen on PRA records even when the spacecraft was far away from Saturn: arcs were very clearly observed as early as July 1981, when Voyager 1 was still at 1 AU. From that distance, the arcs appear more pronounced than when observed closer to the planet (probably because the amplitude of modulation is larger) and their shape is slightly different, the vertex being at lower frequencies (Fig. 2*c*).

The few arcs which are clearly recognized during each rotation are generally at quite different SLS positions. We can study their occurrence in SLS by adding up several rotations. Figure 3 presents two histograms of the position of the arcs for the pre- and post-encounter periods. For each period, a particular sense of arc dominates: VEA before encounter and VLA after encounter, irrespective of longitude.

Arcs of opposite curvature are also observed superposed on the dominant variety, but only in the range of longitudes where the total emission is most intense. Hence, the sub-spacecraft longitudes of these arcs change after encounter. Saturn emission is, at least partly, circularly polarized, but the arc structure does not seem to affect this polarization. We conclude that the sense of polarization remains constant before and after encounter and that the two families of arcs that we recognize during one rotation have the same sense of polarization. A more detailed analysis is required to see whether the arc structure corresponds to a modulation of the degree of polarization.

The difference in the dominant sense of curvature of the arcs between pre- and post-encounter periods can be interpreted in two ways:

(1) The difference could be due to a latitudinal effect because the average latitudes of the spacecraft were different during the inbound (~8°) and outbound (~26°) paths; but whatever the origin of the arc mechanism, it would be more natural to expect a change in the shape of the arcs rather than in their sense of curvature.

(2) The change in the sense of the arcs is a local time effect, a source at fixed position in local time (as invoked to explain the pre- and post-encounter change of the occurrence probability histograms of the emission with sub-Voyager SLS) being seen from two nearly opposite directions before and after encounter. As a local time explanation is more natural for the occurrence histograms, it is also more likely for the change of arc direction.

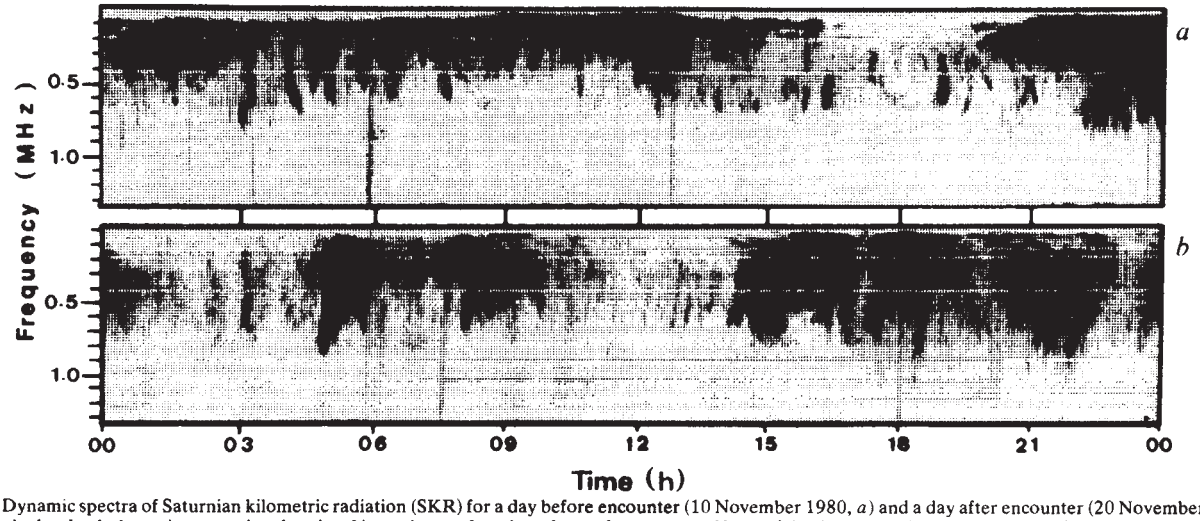


Fig. 1 Dynamic spectra of Saturnian kilometric radiation (SKR) for a day before encounter (10 November 1980, *a*) and a day after encounter (20 November 1980, *b*). In each plot the darkness is proportional to signal intensity as a function of wave frequency and hour of the day. More than two rotations of the planet are covered. The arc structure and the reversal of the direction of the dominant family are clear.

0028-0836/81/340727—02$01.00

© 1981 Macmillan Journals Ltd

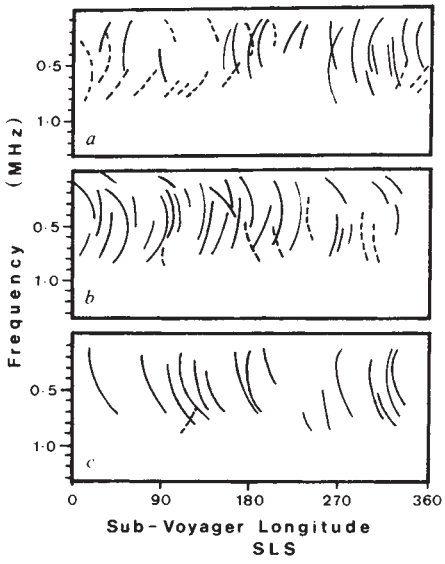


Fig. 2 Arcs observed before encounter (*a*) and after encounter (*b*) for different sub-Voyager longitudes, on different rotations of the planet. *c*, Arcs observed in July and August 1980 when Voyager 1 was ~1 AU from Saturn.

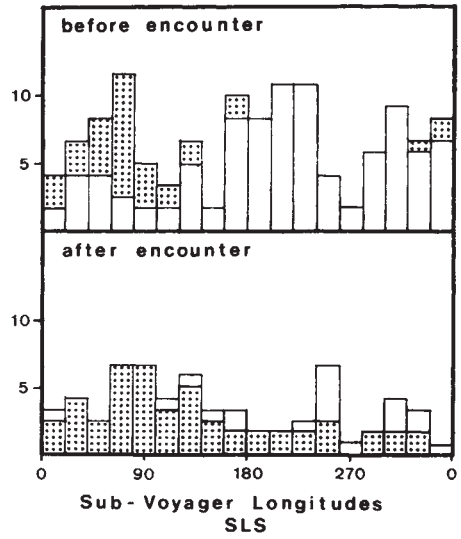


Fig. 3 Histogram of the arc-direction as a function of sub-Voyager longitude in SLS. White corresponds to vertex early arcs (VEA) and dark to vertex late arcs (VLA). At every SLS, VEA dominate before encounter and VLA after encounter, while arcs of the opposite direction are more localized in longitude. Because of the longer period studied before encounter more arcs were seen during this period.

Comparison with Jupiter

The arc structure in Jupiter appears mainly at frequencies larger than 1 MHz. At lower frequencies, the arcs are more difficult to follow, although many drifting features can be seen at the end of the low-frequency leg of the arcs, with their vertex between 5 and 10 MHz. On the other hand, Saturn's emission is observed only at frequencies lower than 1.2 MHz. This large difference in the frequency range of the emission must be considered when we compare the arc structure of the two planets.

The most important result is that arcs cover the whole frequency range of the emission from both Jupiter and Saturn. As the high frequency limit of Saturn's emission is fairly constant the arc shape is stable. In contrast the jovian arc shapes vary considerably with LCM. In particular the higher the frequency limit, the higher the frequency of the vertex. The inclination of the magnetic axis to the rotation axis is much more important for Jupiter than Saturn and explains the difference. As the jovian arc pattern varies consistently with the LCM, we consider that the variation of shape with LCM is principally due to the wobbling of the observer's position in magnetic coordinates due to this inclination. Such a variation does not exist on Saturn because the magnetic and rotation axis nearly coincide[9,10].

Received 13 May; accepted 15 July 1981.

1. Warwick, J. W. *et al. Science* **204**, 995–998 (1979).
2. Leblanc, Y. *J. geophys. Res.* **86** (in the press).
3. Staelin, D. H. *J. geophys. Res.* **86** (in the press).
4. Goldstein, M. L. & Thieman, J. R. *J. geophys. Res.* **86** (in the press).

Another comparison concerns the sense of the arcs and its change at encounter. If we compare the arcs of Saturn with the lesser arcs of Jupiter[1-2], whose emission is not controlled by Io, the variations are similar. The lesser arcs are predominantly VLA before encounter and VEA after encounter, but both types of arc are always present. Hence, although the character of the dominant family is opposite, the reversal at encounter is observed for both planets. Thus the main characteristics of the arcs in Jupiter's and Saturn's emission are similar, even if their frequency ranges are different.

At Jupiter, several types of arcs have been identified[2]: at least two types of lesser arcs, and several types of Io-controlled and non-Io-controlled high frequency arcs, while only two are observed for Saturn. This probably requires several distinct sources of jovian emission while only one or two are necessary to explain the Saturn's arcs. The larger variability of the arc with longitude observed for Jupiter can be explained simply by the change in the magnetic latitude of the observer during one rotation due to the inclination of the magnetic axis to the rotation axis. Therefore the cause of the arcs is probably related to the magnetic field and the similarities between Saturn's and Jupiter's arc structures mean that they are probably due to the same physical mechanism.

5. Boischot, A. & Aubier, M. *J. geophys. Res.* **86** (in the press).
6. Warwick, J. W. *J. geophys. Res.* **86** (in the press).
7. Lecacheux, A., Meyer-Vernet, N. & Daigne, G. *Astr. Astrophys.* (in the press).
8. Warwick, J. W. *et al. Science* **212**, 239–243 (1981).
9. Smith, E. J. *et al. Science* **207**, 407 (1980).
10. Acuña, M. H. & Ness, N. F. *Science* **207**, 444–446 (1980).

Arcs in Saturn's radio spectra

J. R. Thieman* & M. L. Goldstein†

*Advanced Systems Planning Division, ORI Inc, 1400 Spring St Silver Spring, Maryland 20910, USA
†Code 692, Laboratory for Extraterrestrial Physics, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Arcs appearing in the dynamic spectra of Saturn radio emission have been modelled using the same techniques as applied to jovian arcs. Their properties suggest a source region several tenths of a Saturn radius above the cloud tops in the late morning local time sector.

ARC-LIKE structures appear sporadically in the dynamic spectra of Saturn radio emission measured by the Voyager 1 planetary radio astronomy (PRA) experiment. Arcs were pervasive in the dynamic spectra of Jupiter emission[1-3] but they are not nearly as prevalent in Saturn data. There are some differences in the properties of the arcs from the two planets, but a model used to fit jovian arcs has also been successfully applied to Saturn arcs. The curvature and frequency range of the Saturn arcs are best fitted by assuming the source is along magnetic field lines with small L values ($L \leqslant 5$).

Figure 1 shows examples of Saturn dynamic spectra and the types of arcs which occur: both vertex-early and vertex-late arcs (appearing as opening and closing parentheses, respectively) are indicated. A search of dynamic spectra between June and December 1980 resulted in a set of 82 arcs but data gaps during this period preclude determination of an exact occurrence rate. Table 1 summarizes the similarities and differences between the catalogued arcs and the arcs seen in jovian dynamic spectra. As Boischot *et al.*[4] have considered some of these characteristics, only some of the entries will now be discussed.

0028–0836/81/340728—04$01.00

© 1981 Macmillan Journals Ltd

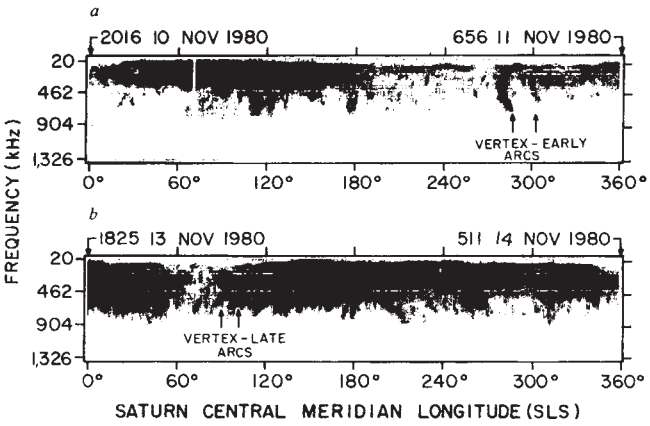


Fig. 1 Examples of arcs in the dynamic spectral plots of Saturn emission detected by the Voyager 1 spacecraft. Data from one rotation of Saturn before (*a*) and after (*b*) closest approach. The beginning and ending spacecraft data acquisition times are marked. Darkness is proportional to the signal strength computed for 1° longitude averages in each of the 70 frequency channels from 1.2 kHz to 1.32 MHz. The abscissa is the longitude of the meridian plane containing the spacecraft. Two vertex-early and two vertex-late arcs are indicated.

Although the right-hand polarization sense (as conventionally used by radio astronomers) is dominant for Saturn arcs, left-hand polarized arcs occur occasionally, mostly during closest approach to Saturn when the spacecraft trajectory was southernmost in latitude[5]. For the 82 arcs analysed, the duration ranged from 10 min to >1 h, with an average of ~45 min.

Figure 2 displays histograms of the minimum, maximum and vertex frequencies for the catalogued Saturn arcs. The average vertex frequency is slightly lower than the value of 400 kHz cited by Boischot *et al.*[4], but their value was based on arcs identified near encounter only. The low frequency extent of the arc peaks sharply around 100 kHz, probably as a consequence of the plasma frequency cutoff due to the plasma density between the source and the spacecraft.

The most interesting difference between Jupiter and Saturn is that for the inbound passage of Voyager 1 to Saturn nearly all of the observed arcs had vertex-early curvature sense while the reverse was true outbound. The sense of curvature of some jovian arcs also shows some pre- and post-encounter dependence, but the dominant influence is longitude. Both vertex-early and vertex-late curvature are often superimposed, and the amount of curvature, the vertex frequency and the frequency extent for each type of curvature are directly dependent on jovian longitude. The inbound and outbound difference in curvature sense at Saturn is interpreted as a local time effect as there is some evidence that the source of emission is confined in local time[6].

Although differing means of arc generation are possible, the many similarities in the planets' arcs suggest similar mechanisms. Beaming of radio emission into a hollow conical sheet was a common hypothesis long before arcs were discovered[7]. Previous modelling work demonstrated that a hollow conical beam which is centred on and rotating with a jovian magnetic field line can reproduce the observed jovian arcs if the cone opening angle is allowed to vary with distance along the field line[8]. To apply this model to Saturn the direction and magnitude of the magnetic field at any point must be known. The model of Saturn's magnetic field which is used is based on Voyager magnetometer experiment results[9]. It can be modified to include the effects of a current sheet[10], but the corrections were not of sufficient magnitude to affect the arc modelling results. The discrete nature of the arcs implies that activity takes place along individual magnetic field lines. For Jupiter the active field lines were usually assumed to have magnetic L values[11] of ~6, as those field lines pass through the orbit of the moon Io which is known to influence the occurrence of radiation. There is no *a priori* reason to limit L values of the field lines for Saturn.

The frequency of the emission is assumed to be just above the right-hand cutoff frequency ω_R so that right-hand polarized emission escapes the planet. The cutoff frequency is defined as:

$$\omega_R = \Omega_e[1+(1+4\omega_{pe}^2/\Omega_e^2)^{1/2}]/2$$

where $\omega_{pe} \equiv \sqrt{(4\pi Ne^2/m)}$ is the electron plasma frequency, $\Omega_e \equiv |eB/mc|$ is the electron gyrofrequency, e, m and N are the electron charge, mass and number density, and B is the magnetic field strength.

The apparent emission sources for the Earth and Jupiter are in underdense plasma regions ($\Omega_e \gg \omega_{pe}$) close to the planets so that ω_R and the emission frequency are very close to the electron gyrofrequency. If we assume that the maximum emission frequency in Saturn radiation corresponds to the maximum gyrofrequency above the planetary cloud tops, then the highest emission frequency would be ~1.5 MHz. Observed emission never exceeds 1.2 MHz and is rarely higher than 900 kHz, so there seems to be a fundamental difference from Jupiter. Low L-value field lines have maximum gyrofrequencies in the 900 kHz range, but the lowest gyrofrequencies on these lines are higher than the lowest radiation frequencies so the source region cannot lie on these field lines. Radio occultation measurements of the Saturn ionosphere show that the plasma frequency close to the planet is comparable with the gyrofrequency[12,13]. The region where the gyrofrequency is ~900 kHz is several tenths of a Saturn radius above the cloud tops (for $L \sim 5$). In this region the plasma frequency is much less than the gyrofrequency (as it was at Jupiter), suggesting that an underdense plasma environment may again be necessary in the source region. Consequently, the maximum frequency in the models of the Saturn arcs has been determined by assuming the emission occurs only beyond a fixed distance from the planetary cloud tops. This is similar to the Earth's auroral kilometric radiation (AKR) which occurs well above the surface at a distance of ~2 Earth radii[14].

The vertex frequency is defined from observations to be a set fraction of the maximum frequency. The minimum arc frequency is difficult to determine since surrounding radiation often mixes with and obscures the low frequency end. The model assumes a fixed lowest frequency at a value determined from observations. The cone angles are allowed to vary with position along the field line from a minimum at the lowest frequency to a maximum at the nose frequency and then back to the minimum

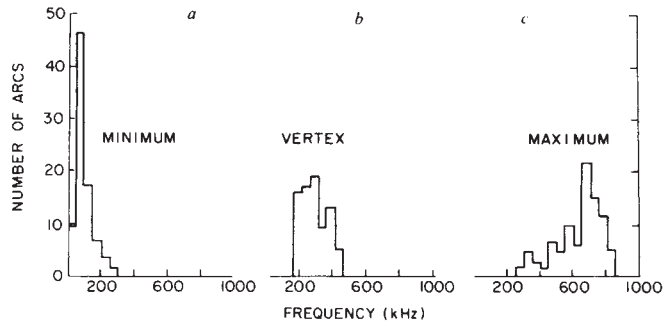


Fig. 2 Histograms of the distribution of the: *a*, minimum; *c*, maximum, and *b*, vertex frequencies in a catalogue of 82 arcs. The catalogue spans 19 June 1980 to 12 December 1980.

Table 1 Saturn–Jupiter arc comparison

	Jupiter	Saturn
Polarization	Dominantly right-hand	Nearly all right-hand
Total arc duration	~20 min	~45 min
Frequency range	~100 kHz to 40 MHz	~50 kHz to ~1.0 MHz
Occurrence rate	Nearly all high band radiation in every rotation made of arcs	At most a few per rotation
Nesting	Large number per group	1–3 per group
Longitude dependence	Curvature and frequency	None
Local time dependence	Curvature	Curvature

© 1981 Nature Publishing Group

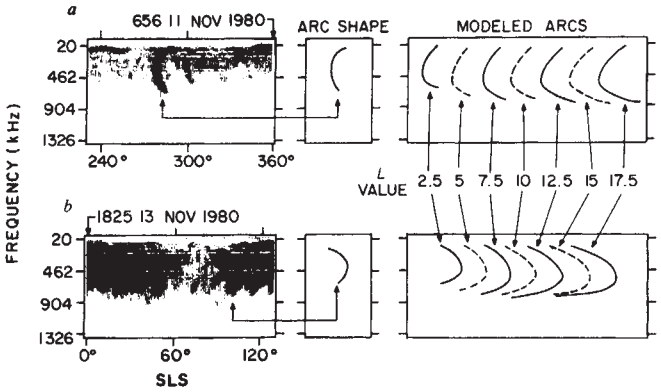


Fig. 3 A comparison of the forms of model arcs to the actual arc shapes derived from the data in Fig. 1. Seven magnetic L values were used to derive the location of seven possible source regions, yielding the arcs shown.

value at the maximum frequency. The formula below is used to vary the cone angle with frequency and is derived from jovian arc studies[8].

$$\psi(f) = \psi_{max} \sin\left(\frac{\pi}{2}\right)\left(\frac{f_{max} - (1-a)f - af_v}{f_{max} - f_v}\right) \qquad f \geq f_v$$

$$\psi(f) = \psi_{max} \sin\left(\frac{\pi}{2}\right)\left[\frac{(1-a)f + af_v - f_{min}}{f_v - f_{min}}\right] \qquad f \leq f_v$$

ψ_{max} and $\psi(f)$ are the maximum cone angle and the cone angle at frequency f. The minimum, maximum and vertex frequencies are represented by f_{min}, f_{max} and f_v and the coefficient a is defined so that $\psi(f) = \psi_{min}$ at both $f = f_{max}$ and $f = f_{min}$.

Arcs from Fig. 1 are compared in Fig. 3 with curves predicted by the arc model. For the vertex-early arcs the model assumes: a cone angle varying from a minimum of 75° to a maximum of 85°; a minimum frequency of 100 kHz; a nose frequency which is 0.6 of the maximum frequency; a minimum distance for the emission region of 0.2 R_s above the cloud tops; and the L values indicated. The same assumptions are used for the vertex-late arcs except for the cone angle which varies from 70° to 85° and the minimum frequency which is increased to 150 kHz. These values should not be considered to be unique as the parameters can vary within a small range without significantly altering the fits. Other arcs differ in shape from examples in Fig. 3 but the parameters can be changed to fit those arcs. The curvature of the modelled arcs is sensitive to the difference between the maximum and minimum cone angles; slight changes in this difference have a noticeable effect. The values of the maximum and minimum cone angles can vary by tens of degrees without noticeably affecting the curvature, provided the difference is held constant. Only values of the cone angle of the order of 60° or larger allow the cone of emission to intersect the line of sight of the Voyager spacecraft. This is consistent with the Jupiter emission model in which large cone angles are necessary to explain the 150° separation in longitude between the two major radio sources, which are believed to originate from opposite sides of the same cone.

Constraining the magnetic L value is an important goal. In general, higher L values cause an increased curvature at high frequencies because of the cone beaming geometry. No severe curvature is evident in the observations, which implies that the emission is confined to $L \lesssim 5$. For Jupiter, the ability to produce good fits assuming a source along the $L = 6$ field line was consistent with the influence of Io on the emission. Four major satellites as well as Saturn's rings are threaded by field lines with L values <5. The resolution of the arc fitting process is not sufficient to distinguish whether any of these objects may be affecting the emission.

The source location implied from the above limits agrees reasonably well with the location deduced by Kaiser *et al.*[6]. That location, however, is confined to a local time region near the noon meridian. For arcs to be formed from hollow conical beams, the beams must rotate with the planet and so intersect the observer's line of sight. If the source is stationary in local time this is not possible. There may, however, be a limited activity region near the noon meridian in which charged particle precipitation and subsequent radio emission is stimulated along a field line as it rotates into the region, continues as the field line rotates, and then dissipates.

The confinement to a specific local time domain explains the dominance of vertex-early arcs in inbound observations and vertex-late arcs after encounter. Figure 4 displays the proposed emission geometry. Inbound, Voyager 1 approached along a local time line of ~13 h. Consequently, the leading edge of the conical beams would usually be the only emission detected. According to the model vertex-early arcs result from the leading edge of a cone, so this is consistent. The outbound trajectory of Voyager 1 was limited to local times around 3 h; thus the trailing edge of the cone was observed and, therefore, vertex-late arcs.

The discrete nature of the arcs for both Jupiter and Saturn remains a problem. For Jupiter, the arcs are discrete but quite regular and usually nested. Saturn arcs appear intermittently

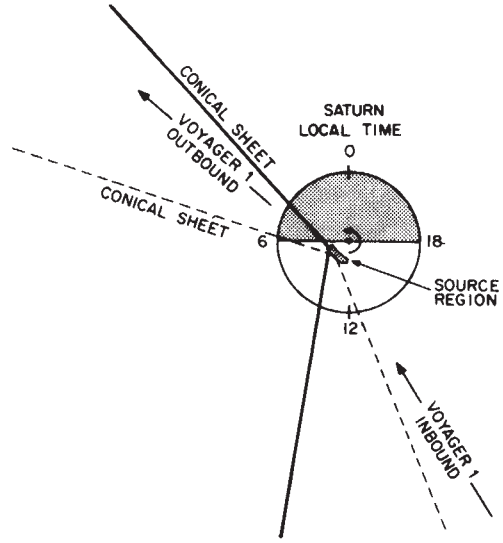


Fig. 4 Hollow conical beam emission from Saturn projected onto the Saturn orbital plane. The source region is confined to the late morning local time sector. Conical emission begins for a particular field line rotating into this region, and ends when the field line rotates out of the region. The directions of the inbound and outbound trajectories are indicated by the arrows. The leading (trailing) edge of the cone rotates through the inbound (outbound) Voyager 1 line of sight.

and are only occasionally nested. Gurnett and Goertz[15] suggested a mechanism for jovian nested arcs which would produce sources along regularly spaced field lines threading Io's plasma torus. Lecacheux *et al.*[16] proposed the production of discrete arcs by a Fresnel screen interference pattern resulting from the observation of the source region through the Io plasma torus. Because there is no similar plasma torus in the magnetosphere of Saturn these theories require modification to explain the discrete Saturn arcs. It may be necessary to look to the solar wind, moons, rings or other influences as sources of discrete excitation.

Our study suggests that the hollow conical emission model of Jupiter arcs can be adapted to fit the Saturn arcs and indicates (1) source regions located along magnetic field lines with $L \lesssim 5$, and (2) a source location several tenths of a Saturn radius above the cloud tops along high latitude field lines. The pre- and post-encounter vertex sense of curvature is consistent with conical emission rotating through the late morning local time region.

The work of J.R.T. was supported by NASA PO S 03389-6C.

Received 13 May; accepted 23 July 1981.

1. Warwick, J. W. *et al. Science* **204**, 995–998 (1979).
2. Warwick, J. W. *et al. Science* **206**, 991–995 (1979).
3. Alexander, J. K., Carr, T. D., Thieman, J. R., Schauble, J. J. & Riddle, A. C. *J. geophys. Res.* (in the press).
4. Boischot, A., Leblanc, Y., Lecacheux, A., Pedersen, B. M. & Kaiser, M. L. *Nature* **292**, 727–728 (1981).

© 1981 Nature Publishing Group

5. Warwick, J. W. *et al. Science* **212**, 239–243 (1981).
6. Kaiser, M. L., Desch, M. D. & Lecacheux, A. *Nature* **292**, 731–733 (1981).
7. Dulk, G. A. *Icarus* **7**, 173–182 (1967).
8. Goldstein, M. L. & Thieman, J. R. *J. geophys. Res.* **86** (in the press).
9. Ness, N. F. *et al. Science* **212**, 211–217 (1981).
10. Connerney, J. E. P., Acuña, M. H. & Ness, N. F. *Nature* **292**, 724–726 (1981).
11. McIlwain, C. E. *J. geophys. Res.* **66**, 3681–3691 (1961).
12. Kliore, A. J. *et al. Science* **207**, 446–449 (1980).
13. Tyler, G. L. *et al. Science* **212**, 201–206 (1981).
14. Alexander, J. K., Kaiser, M. L. & Rodriguez, P. *J. geophys. Res.* **84**, 2619–2629 (1979).
15. Gurnett, D. A. & Goertz, C. K. *J. geophys. Res.* (in the press).
16. Lecacheux, A., Meyer-Vernet, N. & Daigne, G. *Astr. Astrophys.* (in the press).

Saturnian kilometric radiation: statistical properties and beam geometry

M. L. Kaiser*, M. D. Desch* & A. Lecacheux†

* NASA/Goddard Space Flight Center, Laboratory for Extraterrestrial Physics, Greenbelt, Maryland 20771, USA
† Observatoire de Paris, Section d'Astrophysique, F-92190 Meudon, France

Analysis of the average properties of Saturn's kilometre wavelength radio emission suggests that the source region is near the noon meridian in the northern auroral zone and/or the polar cusp.

THE detection[1] of kilometre-wavelength radio emissions from Saturn (SKR) increases the possibility that useful comparative radio 'planetology' can be done, to provide insight into the similarities and differences in particle acceleration and precipitation processes on the three known non-thermal radio planets—Earth, Jupiter and Saturn. However, before Saturn's radio emissions can be studied, an adequate description of their average statistical properties is needed. We now present these properties and discuss the implications for source location, beaming, and magnetic surface anomalies. Such statistical properties are already well known for the Earth's auroral kilometric radiation (AKR). Jupiter's decametric wavelength radio emissions have been studied with ground-based instruments for 25 yrs, and space-borne radio telescopes have extended studies of the jovian spectrum to 10 kHz.

The statistical properties of AKR are a strong function of the observer's viewing geometry. Gurnett[2] first showed that the observer's sub-spacecraft magnetic latitude and local time were the important coordinates in which to organize studies of the Earth's emission beam properties. The Voyager observations of Jupiter have shown that these same two coordinates are of prime importance over much of that planet's radio spectrum[3–6]. Figure 1 shows the magnetic latitude–local time coverage of the three radio planets by the Voyager spacecraft and ground-based observations. For the Earth, >90% of the possible viewing locations have been surveyed frequently enough to provide good statistics; only a small region above the southern polar cap has not been surveyed. For Jupiter, the ground-based observations and the two Voyager fly-bys have covered only ~4% of the possible observing plane, and for Saturn the total coverage will be <0.5% even after both Voyager encounters. This is largely due to the near perfect alignment of the Saturnian magnetic moment with the rotational axis[7–9] which almost eliminates the 'rocking' in magnetic latitude experienced by both the Earth and Jupiter. We describe the average properties of SKR as observed from locations A and B in Fig. 1 by the Voyager 1 planetary radio astronomy (PRA) instrument[10,11] for a 2-month period centred on the 12 November 1980 encounter. Location A specifies the inbound (pre-encounter) geometry (at ~12.5 LT and +8° lat) and location B, the outbound (post-encounter) geometry (at ~3.5 LT and +26°). Location C will be the Voyager 2 outbound trajectory in late 1981. The Saturn hemisphere from 13 to 01 h is only briefly entered by the Voyagers, specifically during their fast swing-bys at closest approach. Additional radio noise from a source in this region would not be detected by the Voyagers.

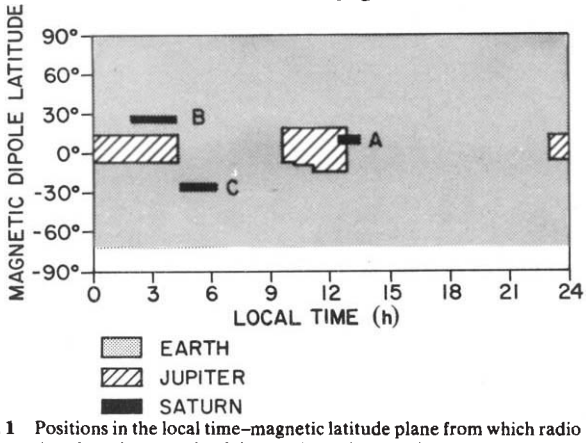


Fig. 1 Positions in the local time–magnetic latitude plane from which radio observations have been made of the Earth, Jupiter and Saturn. Local time is defined as the equatorial plane projection of the observer–planet–sun angle with 12 h = local noon = 0°. A, The inbound trajectories of both voyagers; B, the outbound Voyager 1 trajectory: C, outbound Voyager 2. Thin lines connect A and B, and A and C but they correspond to such brief durations that statistical properties cannot be determined for those viewing locations.

Table 1 Total power of SKR

Occurrence Probability	50%	10%	1%
Inbound power (W)	2×10^8	3×10^9	3×10^{10}
Outbound power (W)	6×10^7	1×10^9	8×10^9

Occurrence probability

An analysis of the occurrence pattern of SKR as a function of Saturn's rotation phase has shown[12] that SKR occurrence is not continuous, but variable and roughly periodic. From pre-encounter observations, a Saturn longitude system[12] (SLS) can be defined. SKR was found to be well organized in longitude with an occurrence rate maximum near 90° and a minimum at ~270° in sub-Voyager SLS. The PRA observations near closest approach[13] showed that the SKR occurrence pattern was fixed relative to the Sun with maximum occurrence rate near 100° in sub-solar SLS.

Figure 2 shows the absolute occurrence rate of SKR bursts stronger than 10^6 Jy (1 Jy = 10^{-26} W M^{-2} Hz^{-1}) as a function not only of sub-solar SLS, but also of observing frequency. (We have adopted 1 AU as a 'standard' distance for normalizing flux density measurements.) The reported[13] similarity in the occurrence patterns for both the inbound and outbound trajectories is clear. However, the absolute occurrence rate for the outbound observations is roughly half that of the inbound observations. Furthermore, the maximization of occurrence in the 100–200 kHz frequency range which is so evident in the inbound panel is replaced in the outbound panel with a broad maximum from 100 to 500 kHz. At the 5% occurrence level, SKR extends from ~ 40 to 600 kHz. SKR events extending to 1,200 kHz have been reported[13], but the occurrence rate of these events (>10^6 Jy) is ≪1%, with no discernable difference between inbound and outbound observations. SKR events as low as 3 kHz have also been reported[14].

Figure 3 shows, for an observing frequency of 174 kHz, the probability of detecting emission above a given flux level for both inbound and outbound trajectories. The probability of

0028-0836/81/340731—03$01.00
© 1981 Macmillan Journals Ltd

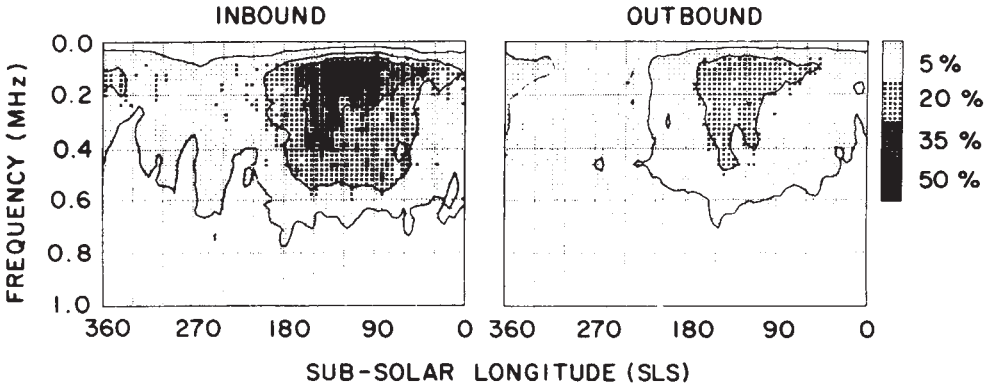


Fig. 2 The absolute occurrence rate of SKR above 10^6 Jy is shown as a function of sub-polar SLS and observing frequency for the Voyager 1 inbound and outbound observing positions (corresponding to A and B in Fig. 1).

detecting a given flux density event is generally higher on the inbound leg than outbound. For example, the probability of detecting SKR stronger than 10^5 Jy is ~58% inbound, but only 32% outbound. For events stronger than 10^6 Jy, the probabilities are 18% inbound, and only 8% outbound.

Several authors[12,13] have reported that SKR is largely right-hand (RH) polarized with left-hand (LH) SKR detected only during the approximately 2-month period immediately preceding encounter, and not at all after encounter. Figure 3 shows the occurrence probability of LH SKR. Its rate of occurrence is only a small fraction of the total SKR occurrence rate. The distribution of LH SKR in frequency and SLS is not well determined due to the sparsity of data, but there is evidence that the LH emission maximizes near 350 kHz at about 300° SLS. This is ~200 kHz higher in frequency and 180° displaced in longitude from the RH polarized emission observed inbound.

Power flux density

Figure 4 shows the average power flux density spectra of SKR at three different occurrence levels for both the inbound and outbound observations. The spectra have been averaged over all longitudes. The general trends evidenced in the occurrence probability distributions (Figs 2 and 3) can be seen. The flux densities observed during the outbound trajectory are lower at frequencies below 500 kHz than the corresponding values inbound. The equivalent isotropic powers for the three occurrence levels are shown in Table 1. Roughly three to four times as much power is received at observing position A than at B.

The frequency of peak flux for inbound observations is near 150–175 kHz. The peak outbound appears to be shifted to higher frequencies, in some cases as high as 400 kHz. However, comparison of the ratio of the corresponding inbound and outbound spectra reveals that this apparent shift towards higher frequencies is caused by a systematic decrease in observed intensities at lower frequencies for outbound observations. For example, for the 10% occurrence spectra, the received flux at 100 kHz is five times larger for the inbound case than for outbound. This ratio decreases to a factor of 2.5 at 250 kHz and approximately equal levels are observed at or above 500 kHz.

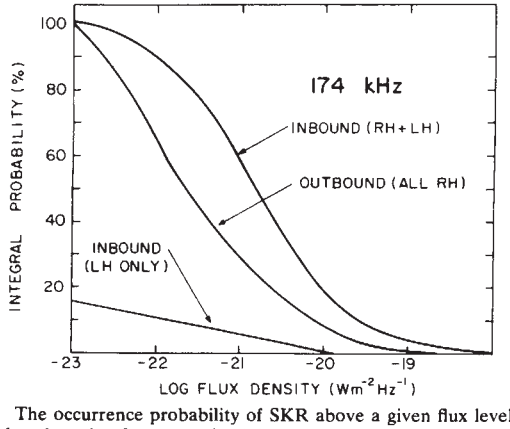


Fig. 3 The occurrence probability of SKR above a given flux level for a particular observing frequency for inbound, outbound and LH polarized events. The flux densities are normalized to a standard distance of 1.0 AU.

Discussion

The statistical SKR properties discussed above strongly constrain possible source locations. We know that SKR emanates from a source which is fixed relative to the Sun, perhaps much like the Earth's AKR. However, SKR would need to be emitted into an extremely sharp beam for the very small magnetic dipole tilt to account for the very large (more than a factor of 100) intensity modulation as a function of rotation[13]. Instead, we favour the concept that emission is generated when a particular longitude arrives at a particular solar hour angle (local time). From the flux level differences inbound and outbound, the radio beam is clearly oriented more towards the inbound trajectory. This is reinforced by the fact that the low frequency portion of each outbound SKR spectrum is depressed. Studies of the Earth's AKR angular distribution[14] show that high frequencies

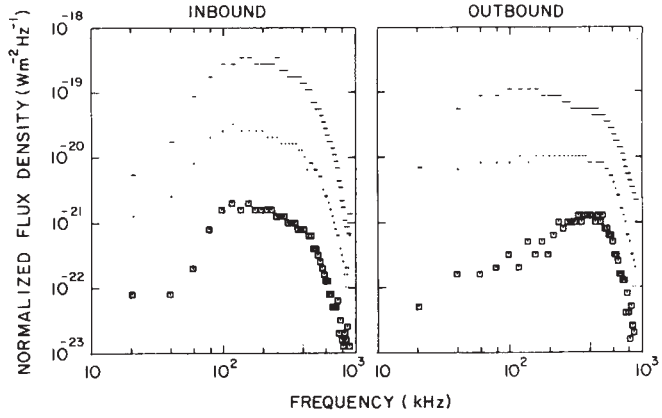


Fig. 4 Typical SKR flux density spectra observed 50% (□), 10% (•) and only 1% (bars) of the time are shown for inbound and outbound observations.

are beamed over a much wider angle than low frequencies due to refraction. If analogy with AKR is correct, a beam oriented more towards the inbound trajectory would tend to illuminate the outbound trajectory at high frequencies, but perhaps not at low frequencies.

From consideration of the inbound spacecraft trajectory and the polarization of SKR, Warwick *et al.*[13] concluded that the RH emission is emitted in the magnetoionic extraordinary mode from the northern hemisphere, and the LH emission is most likely from the southern hemisphere. This implies that SKR is most likely generated near the electron gyrofrequency in the source region. The magnetic field model[7–9], coupled with the observed SKR frequency range, <20–~1200 kHz, suggests that the emission source region is confined to field lines exiting the planet at rather high latitudes, probably >60°. At lower latitudes (<56°) the gyrofrequency does not extend as low as 20 kHz anywhere on the field line.

Additionally, because the SKR comes from a source fixed in local time, the observation[13] that the sense of polarization remains RH inbound and outbound implies that both trajectories are illuminated by the forward lobe of the same beam. In other words, the initial wave direction must be predominantly

©1981Nature Publishing Group

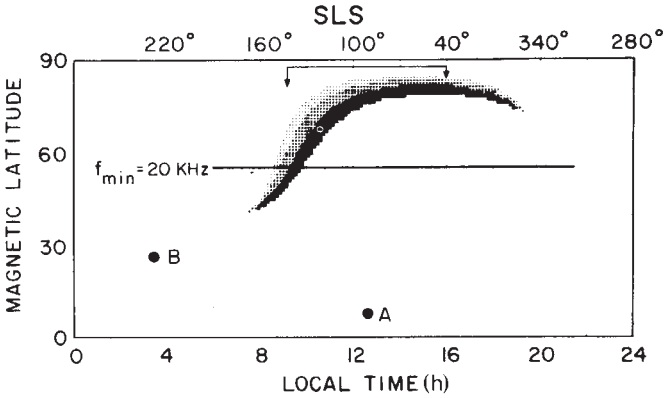


Fig. 5 The shaded region shows, in the saturnian local time–magnetic latitude plane, the locus of possible emission cone axes for frequencies near 175 kHz. The dark shading is for an assumed $\cos^{1/2}\theta$ beam pattern where θ is the angle between the ray and field line, the moderate shading for a $\cos\theta$ pattern, and the lightest shading for a $\cos^2\theta$ pattern. Also shown is a line which corresponds to an equatorial electron gyrofrequency of 20 kHz. Only regions at higher latitudes than those indicated will have emission as low or lower than 20 kHz. The bracketed arrows show the region where similar plots for 20 kHz and 500 KHz intersect this plot. Because the SKR beam appears to be fixed in local time, we can approximately equate longitudes to local time and this scale is shown at the top. A similar analysis done with the cone axis parallel to the field gradient direction gave essentially the same result.

parallel to Saturn's magnetic field for both the A and B viewing locations. Thus, the SKR RH beam seems to emanate from the northern hemisphere in the sector containing A and B, and oriented more towards A.

We can speculate on the source location by making some assumptions concerning the shape and orientation of the SKR beam. The Earth's AKR beam near the frequency of peak flux density is very broad, closely resembling a cosine pattern with the beam axis oriented more or less parallel to the field[15]. If we assume that SKR is also emitted in a broad, symmetrical beam pattern, and that the axis of the pattern is also oriented parallel to the field, we can determine where the axis must be in local time and latitude so that intensities received at A exceed those at B by some factor consistent with Table 1. Figure 5 shows a plot similar to Fig. 1 indicating the locus of possible beam axes for three different broad beams where A would receive 2–4 times as much flux as B. The calculations were made at altitudes corresponding to 175-kHz electron gyrofrequency. The field lines yielding solutions were then traced to the surface (invariant latitude). Several source locations are possible, but most intriguing is the region at high latitudes near the noon meridian. This region most likely corresponds to the auroral zone[9,16], and, by analogy with the Earth[2] and Jupiter[6], is a very good candidate for a region of enhanced particle precipitation required to produce radio emission. Additionally, the apparent confinement of the radio source to a limited dayside range of local times suggests an association with a dayside polar cusp in Saturn's magnetosphere[17]. Because SKR is fixed in local time, we can equate local time and sub-solar SLS as shown at the top of Fig. 5. The SLS corresponding to the proposed source region (40°–145°) may correspond to a region of anomalous magnetic field, although it is not clear that a precipitation producing field anomaly need be in the source region itself. The Voyager 1 close approach to Saturn was at high southern latitudes and displaced some 90–100° in longitude from the proposed source location, so fields and particle experiments were not able to make measurements on field lines threading this region. However, the Voyager 2 close approach on 25 August 1981 will be in a much better position to make measurements over this region.

Received 13 May; accepted 15 July 1981.

1. Kaiser, M. L. *et al. Science* **209**, 1238–1240 (1980).
2. Gurnett, D. A. *J. geophys. Res.* **79**, 4227–4238 (1974).
3. Alexander, J. K. *et al. J. geophys. Res.* **84**, 5167–5174 (1979).
4. Alexander, J. K. *et al. J. geophys. Res.* **86** (in the press).
5. Desch, M. D. & Kaiser, M. L. *J. geophys. Res.* **85**, 4248–4256 (1980).
6. Boischot, A. *et al. J. geophys. Res.* **86** (in the press).
7. Smith, E. J. *et al. Science* **207**, 407–411 (1980).
8. Acuña, M. H. & Ness, N. F. *Science* **207**, 444–446 (1980).
9. Ness, N. F. *et al. Science* **212**, 211–271 (1981).
10. Lang, G. J. & Peltzer, R. G. *IEEE Trans.* **AES-13**, 466–471 (1977).
11. Warwick, J. W. *et al. Space Sci. Rev.* **21**, 309–319 (1977).
12. Desch, M. D. & Kaiser, M. L. *Geophys. Res. Lett.* **8**, 253–256 (1981).
13. Warwick, J. W. *et al. Science* **212**, 239–243 (1981).
14. Gurnett, D. A. *et al. Science* **212**, 235–239 (1981).
15. Green, J. L. *et al. J. geophys. Res.* **82**, 1825–1838 (1977).
16. Broadfoot, A. L. *et al. Science* **212**, 206–211 (1981).
17. Behannon, K. W. *et al. Nature* **292**, 753–756 (1981).

Narrowband electromagnetic emissions from Saturn's magnetosphere

D. A. Gurnett*, W. S. Kurth* & F. L. Scarf†

*Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52242, USA
†TRW Defense and Space Systems, One Space Park, Redondo Beach, California 90278, USA

A series of narrowband electromagnetic emissions were detected by the plasma wave instrument on board Voyager 1 coming from the inner region of Saturn's magnetosphere in the frequency range 3–30 kHz. These emissions have many similarities to continuum radiation detected in the Earth's magnetosphere and narrowband kilometric radiation in the jovian magnetosphere. The observed frequency spacing suggests that the emissions are being generated near Tethys, Dione and Rhea, probably in regions of large plasma density gradients associated with boundaries of the plasma sheet.

During the Saturn flyby the Voyager 1 plasma wave instrument observed a series of narrowband electromagnetic emissions at frequencies between 3 and 30 kHz (ref. 1). These bands are remarkable because of their extremely complex spectral structure, consisting of a large number of nearly monochromatic emissions with frequency spacings corresponding to the equatorial electron gyrofrequencies near the moons Tethys, Dione and Rhea. Because of the possible association of the radio emissions with the moons of Saturn, we have analysed these emissions to determine their origin and to explore similarities to other planetary radio emissions. We describe here their characteristic and discuss models of the generation of these emissions (the plasma wave instrumentation is described in ref. 2).

Observations

The strongest and best resolved examples of the narrowband electromagnetic emissions occur in the inner region of the magnetosphere, near closest approach. The electric field intensities in this region are shown in Fig. 1. Each channel gives the electric field intensity on a logarithmic frequency scale at the centre frequencies indicated on the left of the plot. The electron gyrofrequency, f_g, and plasma frequency, f_p, are also shown for reference. The electron gyrofrequency ($f_g = 28B$ Hz, where B is the magnetic field in nT) was determined from the Voyager 1 magnetic field instrument[3], and the plasma frequency ($f_p = 9{,}000\sqrt{n_e}$ Hz, where n_e is the electron number density cm^{-3}) was obtained from the Voyager 1 plasma instrument[4]. The narrow-

0028–0836/81/340733—05$01.00

© 1981 Macmillan Journals Ltd

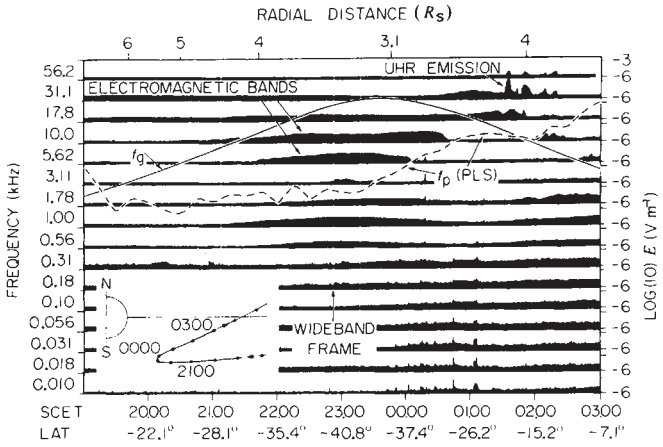


Fig. 1 The plasma wave electric field intensities in the region near closest approach. The narrowband electromagnetic emissions are confined to the low density region outside of the plasma sheet at $f > f_p$ (dashed lines). The plasma sheet is entered at about 00.00 to 01.00, as the spacecraft approaches the magnetic equator (see trajectory sketch).

band electromagnetic emissions consist of a series of smooth enhancements in the 3.11–31.1-kHz channels from about 21 h 00 min (spacecraft event time) on 12 November to 01 h 00 min on 13 November. These enhancements all occur at frequencies above the electron plasma frequency as determined by the plasma experiment. The sharp cutoffs evident in the 5.62- and 10.0-kHz channels shortly after closest approach seem to be directly associated with the local electron plasma frequency. The enhanced intensities extend across the electron gyrofrequency with no obvious propagation effect. These characteristics identify the mode of propagation as the free-space left-hand polarized ordinary $(L, 0)$ mode. For this mode propagation occurs for all frequencies $f > f_p$, and that no resonance or cutoff occurs at the electron gyrofrequency.

Another type of very intense narrowband emission occurs in the 31.1-kHz channel of Fig. 1 at about 01.30 on 13 November, coincident with the cutoff of the electromagnetic radiation. At peak intensity this narrowband burst has a field strength ~60 db larger than the adjacent radiation. Based on similar observations in the Earth's magnetosphere[5–7] and at Jupiter[8–10] this burst is identified as an electrostatic emission at the local upper-hybrid resonance (UHR) frequency, $f_{UHR} = (f_p^2 + f_g^2)^{1/2}$. The fact that the observed emission frequency is higher than the upper hybrid resonance frequency computed from the measured plasma frequency and gyrofrequency is attributed to uncertainties in the determination of the electron number density. The electron plasma frequency is a factor of ~3 higher than given by the plasma instrument at this particular time.

Because the intensity of the electromagnetic radiation is comparable in adjacent frequency channels, Fig. 1 implies that the spectrum is nearly continuous, possibly comparable with the continuum radiation detected in the Earth's magnetosphere[5]. Fortunately, one 48-s frame of wideband waveform data obtained in this region provides greatly improved frequency resolution. The high-resolution spectrum of the wideband data (Fig. 2) is not continuous, but rather consists of many narrowband emissions, some with bandwidths < 100 Hz. Three main bands can be identified at frequencies of ~1 kHz, ~6 kHz, and ~9.7 kHz. Because the low-frequency band at 1 kHz is below the electron plasma frequency, whereas the others are all above it, this band cannot be propagating in the same mode as the higher frequency bands. For $f < f_p$ the only possible modes are the whistler-mode and the z-mode. Based on the similarity to whistler-mode hiss emissions observed in the higher density regions of the plasma sheet, we concluded that this low-frequency (~1 kHz) band is propagating in the whistler mode[1].

The most definitive evidence that the high-frequency bands are propagating in the free-space electromagnetic mode is given in Fig. 3, which shows a series of wideband spectrograms obtained over a 3-day period near and after closest approach. A single persistent band of emission can be seen at ~5 kHz extending over a range of radial distances from at least 3.26 to 58.3 R_S. Because the plasma parameters change over a wide range it is difficult to see how this band of emission could be observed over such a large range of radial distance without being a freely propagating electromagnetic wave. The last observation, Fig. 3 F, is in the magnetosheath where the local plasma frequency is ~3.8 kHz [J. Scudder, personal communication]. Although the electromagnetic bands can be detected in the wideband data over several days, the intensity and detectability vary considerably. The meridian plane plot in Fig. 4 summarizes the regions in which the bands were detected. Except for the brief interval near closest approach all of the emissions were confined to high latitudes on the outbound pass. None were found on the inbound leg and this is almost certainly due to the fact that the spacecraft was passing through the high density central region of the plasma sheet where the electron density is high enough to refract the radiation away from the equatorial region. In contrast, for the outbound pass ($R \gtrsim 10\ R_S$) at high latitudes the electron density[4] is always low enough to allow direct line-of-sight propagation to the spacecraft at frequencies above a few kHz. The strongest emissions were observed at position A near closest approach, suggesting that the radiation originates in the inner region of the magnetosphere at $R \lesssim 10\ R_S$.

Examination of the 16-channel survey plots shows evidence of a periodic modulation of the narrowband emission intensities at the 10 h 40 min rotation period of Saturn. Typically, the narrowband emission events last for ~3 or 4 h and tend to recur at intervals of ~11 h. The three intervals containing B, C–D, and E, in Fig. 4 show this periodicity. Maximum intensity occurs when the Saturn longitude system (SLS) longitude of the subsatellite point is near 290°. Event F in Fig. 4 is also in phase with this general rotational modulation, however, one cycle seems to have been missed between E and F, possibly owing to the very low intensity and difficulty of detecting these emissions far from the planet.

Relationship to other planetary radio emissions

Although the complex narrowband structure of these Saturn radio emissions appears unusual, very similar types of radio emissions have been identified in the magnetospheres of Earth and Jupiter. At Earth a relatively weak type of broadband radio emission called continuum radiation is generated within the magnetospere[5,11,12]. This has a relatively smooth steady temporal structure, and when analysed with spectral resolution comparable with that of the Voyager 16-channel analyser, the

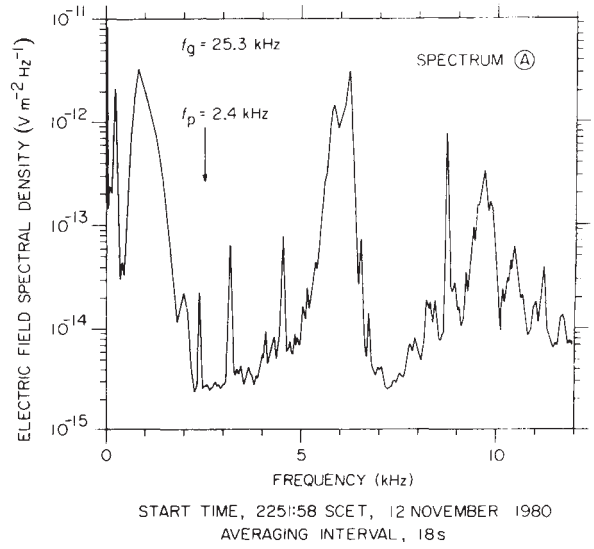


Fig. 2 A detailed frequency spectrum of the emission in Fig. 1, showing the occurrence of numerous bands with approximately harmonic frequency spacings.

© 1981 Nature Publishing Group

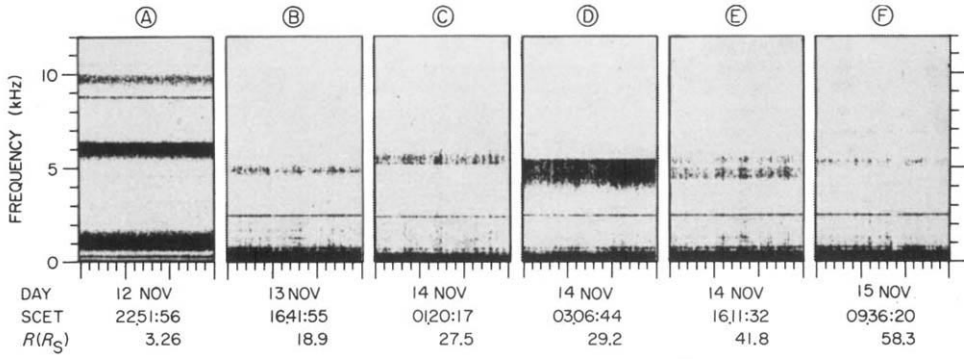


Fig. 3 Frequency–time spectrograms obtained in various regions of the magnetosphere demonstrating that the band at ~5–6 kHz can be detected over a large range of radial distances. This series of observations provides the best evidence that the noise consists of freely propagating electromagnetic radiation.

overall appearance is very similar to the emissions shown in Fig. 1. When the terrestrial continuum radiation is analysed with better frequency resolution it becomes apparent that the spectrum is not continuous, but instead consists of a series of discrete narrowband emissions[13]. The detailed characteristics of the continuum spectrum depend on the frequency range. At frequencies below the solar wind plasma frequency, where the radiation is trapped in the low-density magnetospheric cavity, the spectrum tends to be nearly continuous with only occasional narrowband structure[11]. At frequencies above the solar wind plasma frequency, where the radiation can freely escape from

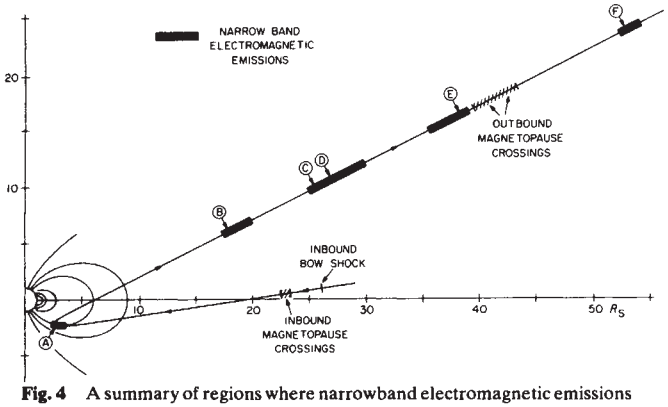


Fig. 4 A summary of regions where narrowband electromagnetic emissions were observed. The black bands indicate detections in the 16-channel spectrum analyser data, and the circles correspond to the wideband spectrograms shown in Fig. 3.

the magnetosphere, the frequency spectrum usually shows considerable narrowband structure. Typically, it consists of several relatively broad emission bands consisting of numerous narrowband features, some of which have harmonic frequency spacings[13]. The spectral characteristics of the terrestrial narrowband emissions are remarkably similar to those observed at Saturn.

At Jupiter narrowband electromagnetic emissions (narrowband kilometric radiation (nKOM)) with similar characteristics have been reported by Kaiser and Desch[14]. Although the frequency resolution available for the jovian nKOM observations is not sufficient to reveal the extraordinary fine structure evident in the terrestrial and saturnian emissions, the overall characteristics are quite similar. In particular, the jovian nKOM intensity is controlled by the rotation of Jupiter, very similar to the rotational control of the narrow-band emissions observed at Saturn. The period of the nKOM is slightly slower than the rotation period of Jupiter, and corresponds to the rotation period of the plasma in the outer region of the Io plasma torus[14], at a radial distance of about 10 R_J.

The close similarities between the narrowband electromagnetic emissions at Earth, Jupiter, and Saturn strongly suggest that the same basic emission mechanism operates at all three planets. At Earth it is now widely believed that continuum radiation is generated by mode conversion from electrostatic waves generated by low-energy electrons. This general mechanism has a long history of experimental and theoretical development[5,7,15,16], including the observation of enhanced radiation intensities when low-energy electrons are injected into the magnetosphere[17]. The electrostatic wave involved is an electrostatic mode which occurs near the upper hybrid resonance frequency, $f_{UHR} = (f_p^2 + f_g^2)^{1/2}$. One of the specific conditions which must be satisfied for this mode to be unstable is that $(n + 1/2)f_g \simeq f_{UHR}$ (refs 6, 7, 18). The physical situation in the Earth's magnetosphere is illustrated in Fig. 5, which shows a representative radial profile of f_{UHR} and f_g near the plasmapause. Intense electrostatic emissions occur whenever f_{UHR} crosses a half-integral harmonic of f_g (dashed lines). Kurth *et al.*[13] have demonstrated that the electrostatic emissions are directly converted to electromagnetic radiation with little or no frequency shift. The radiation is generated in the free space (L, 0) mode. The resulting radio emission spectrum therefore consists of a series of lines with a frequency spacing characteristic of the electron gyrofrequency in the source region. Kurth *et al.*[13] have proposed that the jovian nKOM radiation can be explained by essentially the same mechanism, with the density gradient in the outer region of the Io plasma torus having the same role as the plasmapause at Earth. Large plasma density gradients are an essential element of some theories for the electrostatic to electromagnetic mode conversion[19].

Interpretation of Saturn observations

Because of the close similarities between the narrowband radio emissions at Earth, Jupiter and Saturn, the electrostatic to electromagnetic mode conversion mechanism probably also accounts for the narrowband emissions at Saturn. This view is supported by the fact that intense upper hybrid waves are observed at Saturn in the region where the narrowband electromagnetic radiation is most intense (see Fig. 1). Also, at Saturn these emissions are known to be propagating in the free space (L, 0) mode—the favoured mode for mode conversion. But the exact source location and the physical processes which lead to the generation of intense electrostatic waves remain uncertain.

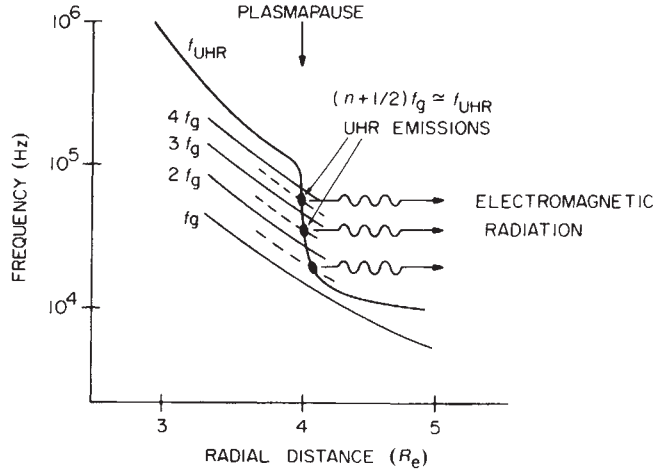


Fig. 5 A model illustrating the mechanism for generating narrowband electromagnetic emissions (continuum radiation) in the Earth's magnetosphere. Intense electrostatic upper hybrid resonance (UHR) emissions at $(n + 1/2)f_g \simeq f_{UHR}$ are believed to be converted to electromagnetic emissions in regions with steep plasma density gradients, such as near the plasmapause.

© 1981 Nature Publishing Group

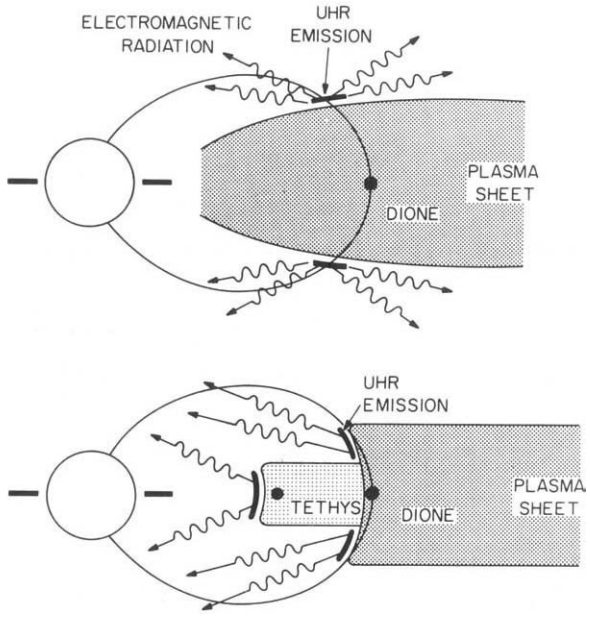


Fig. 6 Two possible models for generating narrowband electromagnetic emissions in Saturn's magnetosphere. *a*, The emissions are generated at the density gradient which exists at the north and south boundaries of the plasma sheet; *b*, the emissions are generated along the field-aligned boundaries of the plasma sheet[4].

Unfortunately, because of no appreciable tilt in the magnetic dipole axis of Saturn, there are no geometry-dependent propagation effects to help determine the source location. Probably the best clue is the frequency spectrum. If the mode conversion model is correct then the frequency spacing between the lines is approximately the electron gyrofrequency in the source region. Figure 2 shows that several half-integral harmonic frequency spacings can be identified in the radio emission spectrum. If we identify the main emissions at 6.0 and 9.7 kHz as a harmonic pair, then the gyrofrequency is 3.7 kHz. A second set of lines can be identified at 3.1, 4.7 and 6.1 kHz, which gives a gyrofrequency of 1.50 kHz. Finally, a third set of lines at 8.7, 9.6, 10.4 and 11.4 kHz gives a gyrofrequency of 870 Hz. Assuming that the source is located near the equatorial plane (as at Earth and Jupiter) the corresponding radial distances for these gyrofrequencies are 5.4, 7.3 and 8.8 R_S, based on a dipole moment of 0.21 G R_S^3 (ref. 3). Note that depending on the plasma density gradient in the source region and the location of the source relative to the magnetic equator, the observed emission frequencies may not occur at exact half-integral harmonics. Therefore, these radial distances should be regarded as only a rough indication of the source position. The radial distances obtained are in the vicinity of Tethys, Dione and Rhea. Because at least one of Saturn's moons, Dione, is already thought to have an important role in controlling Saturn's kilometric radio emissions[1,20,21], this comparison strongly suggests that one or more of these moons is involved in generating the narrowband electromagnetic emissions.

The most likely way a moon could control nonthermal radio emission processes in the inner region of Saturn's magnetosphere is by injecting plasma much as Io controls radio emissions by injecting plasma into the jovian magnetosphere. Evidence from the Pioneer 11 and Voyager 1 plasma instruments[4,22] suggests that plasma production by Tethys and Dione may be the source of the extended disk-shaped plasma sheet observed near the equatorial plane of Saturn. If nonequilibrium processes associated with the plasma injection are responsible for the narrowband electromagnetic emissions, then the analogies with Earth and Jupiter would suggest that the source is associated with a steep plasma density gradient near the edge of the plasma sheet. Figure 6*a*, *b* shows two extreme models for the source positions. In Fig. 6*a* the source is located near the north–south boundaries of the plasma sheet, and in Fig. 6*b* near the sharp field-aligned inner boundary of the plasma sheet described by Bridge *et al.*[4].

If the source is located as in Fig. 6*a* then it would have to be tightly confined to a specific L-shell to produce the narrow emission spectrum. Such a localization could, for example, be caused by a field-aligned current system flowing along the Dione L-shell. A source located on the Dione L-shell at the north–south boundary of the plasma sheet 2 R_S from the equatorial plane would have an electron gyrofrequency in close agreement with the 3.7-kHz major frequency spacing evident in Fig. 2. On the other hand, if the source is located along a field-aligned inner boundary of the plasma sheet then the narrowband characteristic of the source would be easily explained as the gyrofrequency is, to first order, constant along the magnetic field line near the equator. This situation could be closely analogous to the generation of continuum radiation near the plasmapause at the Earth, except that the decreasing density gradient is facing towards rather than away from the planet. Note that the 7.3 R_S radial distance identified for the 1.5 kHz frequency spacing in Fig. 3 closely agrees with field-aligned boundary identified by Bridge *et al.*[4] at $L = 7$. This boundary occurs slightly beyond the orbit of Dione. Because of uncertainties concerning the plasma density distribution in the inner region of the magnetosphere, it is too early to decide which can best account for the characteristics of these emissions.

Conclusion

The Voyager 1 flyby of Saturn has revealed narrowband electromagnetic emisssions originating from the inner region of the magnetosphere. At closest approach the broadband electric field strength of these emissions is $\sim 50\ \mu V\ m^{-1}$, which corresponds to a total radiated power of about 2.5×10^6 W, assuming an isotropic source. Although this is small compared with the power radiated by Saturn at kilometric wavelengths, $\sim 10^8$–10^9 W (ref. 20), these emissions are of interest because of the close similarity to narrowband radio emissions from Earth and Jupiter.

Comparisons with similar emissions in the Earth's magnetosphere suggest that they are produced by mode conversion from electrostatic waves near the upper hybrid resonance. This radio emission mechanism has also been proposed to explain certain types of solar radio emissions[23]. Although the basic mechanism has been identified, the conditions in which the radiation can occur are poorly understood. Observations at all three planets suggest that distinct 'hot spots' occur from which the radiation is preferentially emitted. The association of the source with a plasma density gradient may be a requirement for the mode conversion process[19]. Unfortunately, the requirements for generating the electrostatic waves are not completely understood. The generation of these waves probably involves the presence of large fluxes of low energy (100 eV to 1 keV) electrons with a pronounced loss-cone or $\partial f/\partial v_\perp > 0$ feature and a suitably large cold-to-hot electron density ratio[18,24,25]. The apparent control of the narrowband emission intensities by Saturn's rotation presumably provides an important clue to the processes involved. However, because of the very close alignment of the dipole axis with the rotational axis of Saturn[3], it is not clear how this rotational control is imposed.

Finally, at Saturn no evidence was found for trapped continuum radiation comparable to the continuum radiation trapped in the low density magnetospheric cavity of Earth and Jupiter. As the Voyager 1 plasma measurements[4] demonstrate the existence of a low-density cavity at Saturn with an electron plasma frequency well below the solar wind plasma frequency, it is surprising that no radiation is trapped in this cavity. The radio emission source at Saturn does not seem to extend down to sufficiently low frequencies to illuminate the cavity. The reasons for this marked difference in the radio emission spectrum are not understood.

Work at The University of Iowa was supported by NASA contract 954013 with the Jet Propulsion Laboratory and grants

© 1981 Nature Publishing Group

NGL-16-001-002 and NGL-16-001-043 with NASA Headquarters, and by the Office of Naval Research. Work at TRW was supported by NASA contract 954012 with the Jet Propulsion Laboratory.

Received 13 May; accepted 17 July 1981.

1. Gurnett, D. A., Kurth, W. S. & Scarf, F. L. *Science* **212**, 235–239 (1981).
2. Scarf, F. L. & Gurnett, D. A. *Space Sci. Rev.* **21**, 289–308 (1977).
3. Ness, N. F. *et al. Science* **212**, 211–217 (1981).
4. Bridge, H. S. *et al. Science* **212**, 217–224 (1981).
5. Gurnett, D. A. *J. geophys. Res.* **80**, 2751–2763 (1975).
6. Christiansen, P. *et al. Nature* **272**, 682–686 (1978).
7. Kurth, W. S. *et al. J. geophys. Res.* **84**, 4145–4164 (1979).
8. Scarf, F. L. *et al. Science* **204**, 991–995 (1979).
9. Gurnett, D. A. *et al. Science* **206**, 987–991 (1979).
10. Birmingham, T. J. *et al. J. geophys. Res.* (in the press).
11. Gurnett, D. A. & Shaw, R. R. *J. geophys. Res.* **78**, 8136–8149 (1973).
12. Frankel, M. S. *Radio Sci.* **8**, 991–1005 (1973).
13. Kurth, W. S. *et al. J. geophys. Res.* **86**, 5519–5531 (1981).
14. Kaiser, M. L. & Desch, M. D. *Geophys. Res. Lett.* **7**, 389–392 (1980).
15. Jones, D. *Nature* **260**, 686–689 (1976).
16. Melrose, D. B. *J. geophys. Res.* **86**, 30–36 (1981).
17. Gurnett, D. A. & Frank, L. A. *J. geophys. Res.* **81**, 3875–3885 (1976).
18. Rönnmark, K. *et al. Space Sci. Rev.* **22**, 401–417 (1978).
19. Jones, D. *Nature* **288**, 225–229 (1980); *Adv. Space Res.* **1**, 333–336 (1981).
20. Warwick, J. W. *et al. Science* **212**, 239–243 (1981).
21. Kurth, W. S. *et al. Nature* **292**, 742–745 (1981).
22. Frank, L. A. *et al. J. geophys. Res.* **85**, 5695–5708 (1980).
23. Kuijpers, J. *Radio Physics of the Sun* (eds Kundu, M. & Gergely, T.) 341–361 (Reidel, Dordrecht 1980).
24. Hubbard, R. F. & Birmingham, T. J. *J. geophys. Res.* **83**, 4837–4850 (1978).
25. Ashour-Abdalla, M. & Kennel, C. F. *J. geophys. Res.* **83**, 1531–1543 (1978).

Saturn's radio emissions: rotational modulation

C. K. Goertz*†, M. F. Thomsen*‡ & W.-H. Ip*

*Max-Planck-Institut für Aeronomie, 3411 Katlenburg-Lindau 3, FRG
†Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52240, USA
‡Los Alamos Scientific Laboratory, Los Alamos, New Mexico, 87545, USA

The unexpected rotational modulation of the Saturn kilometric radiation and Saturn electrostatic discharges, as revealed by Voyager 1 observations, are discussed in terms of a ring-current system and the geometry of the dipole field.

THE Saturn flyby observations of Pioneer 11 and Voyager 1 have revealed that the Saturnian ring system has important effects on the structures of the planetary ionosphere as well as the magnetospheric environment, such as the unexpected rotational modulation of the Saturn kilometric radiation (SKR) and the Saturn electrostatic discharge (SED) as discovered by the Voyager planetary radio astronomy experiment[1]. The emissions of the bursty and highly polarized SKR ($f < 1{,}200$ kHz) and the broad banded and unpolarized SED ($f \sim 20.4$ kHz–40.2 MHz) of impulsive nature strongly peaks at the subsolar longitude (SLS) of $\sim 110°$. The rotational control of the SKR and SED is a clock-like effect—with radiations preferentially emitted when Saturn has a particular phase relative to the Sun[1,2]. Although the sources of these radio emissions are not clear, Warwick *et al.*[1] argued that the SED emissions most likely originate in the rings of Saturn.

The challenge then is to devise a mechanism which can produce electrodynamic coupling between the ring system and the planetary ionosphere and/or magnetosphere in which the associated SKR and SED radio emissions are subject to longitudinal control. One obvious candidate is the establishment of field-aligned current systems—connecting the rings and the ionosphere with longitudinal variation. Local acceleration processes in the vicinity of the rings are also needed. Although the solution will probably have to await further observations and more detailed data analyses, we now discuss the interplay between the ring-current systems and the geometry of the dipole field and the rings and describe some of the basic ingredients in the longitudinal modulations of SKR and SED.

The Pioneer 11 radio measurements[3] and the Voyager 1 EUV experimental results[4] reveal that a strong coupling between the neutral hydrogen atmosphere of the rings with a number density of 400–600 cm^{-3} (refs 4–7) and the ionospheric plasma. Assuming the ionospheric H^+ ions have a number density of $n_i \sim 3 \times 10^3\ cm^{-3}$ near the ring plane, the pertinent charge-exchange loss time scale for the neutral H atoms can be estimated to be $\sim 10^6$ s. Thus, for a total volume of the ring atmosphere to be $\sim 3 \times 10^{30}\ cm^3$, the charge-exchange loss rate of the neutral atoms as a result of the $H^+ + H \rightarrow H + H^+$ process is $\dot{N} \sim 10^{27}\ s^{-1}$. This value is perhaps the most direct estimate for the source strength of the ring atmosphere and it supports the assessment by Carlson[8] that photo-sputtering of atomic hydrogen atoms from H_2O ice on the ring particles is responsible for supplying the observed H cloud of the rings, as other proposed effects[9,10] are far less efficient.

Note that 'the acceleration of the new ions (from charge exchange and other ionization effects) to co-rotation with the magnetosphere will necessitate the establishment of a so-called pickup current which flows radially outward outside the synchronous orbit (at 1.8 R_S) but inward inside this distance[11,12]. An order-of-magnitude approximation can be obtained by using $I \sim \dot{M}\Delta v / B\Delta x$ for the total current. Here $\dot{M}$ is the mass addition rate, B is the average magnetic field ($\sim 2{,}500$ nT), $\Delta v \sim 1\ km\ s^{-1}$ is the difference between the corotating speed and the keplerian orbital speed, and Δx is the characteristic radial dimension of such a ring-current system. Even when the ionization rate of oxygen atoms is set equal to that of neutral H atoms, $I \lesssim 10^3$ A for $\Delta x \sim 1\ R_S$. This pickup current system is therefore relatively small, and not subject to rotational modulation.

Another current system is more promising. The near-perfect dipole field of Saturn has a slight vertical displacement (h) of the dipole centre from the planet's centre along the axial direction[13–15]. This enables charged particles to mirror just above the ring plane without being absorbed—a sort of plasma disk may be formed there. The longitudinal drifts (due to magnetic curvature force, gravity etc.) of the ions and electrons in this region would

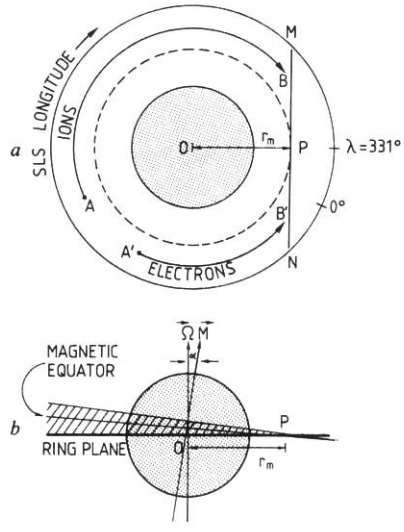


Fig. 1 *a*, Top view of the intersection between saturnian rings and the magnetic equator. The two planes intersect at $\overline{MN}$ and $\overline{OP} = 1.64\ R_S$ if the vertical displacement is 0.02 R_S and the tilt angle α is 0.7°. Inside the synchronous orbit r_{syn} at 1.8 R_S the ions drift west (A → B) and electrons east (A' → B'). Beyond r_{syn} they move in opposite directions. *b*, Side view of the above system. Particles with near 90° equatorial pitch-angle in the wedge-shaped 'free-zone' are not subject to any absorption.

0028–0836/81/340737—03$01.00

© 1981 Macmillan Journals Ltd

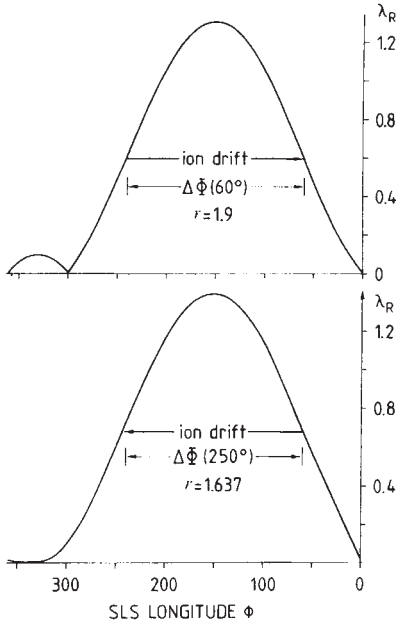


Fig. 2 The magnetic latitudes λ_r of the ring for $h = 0.02$ and $\alpha = 0.7°$ for two radial distances: *a*, $r = 1.9\ R_S$; *b*, $r = 1.64\ R_S$. Also indicated are longitudinal ranges of ions between ionization and ring absorption.

then constitute a current flowing in the azimuthal direction. At the same time, there is also a small tilt (α) of the dipole axis with respect to the spin axis of Saturn towards SLS longitude 331° (ref. 15). The important feature is that the magnetic equatorial plane and the ring plane intersect along a line which has its minimum distance from Saturn (r_m) at an SLS longitude of 331°. For $\alpha = 0.7°$ and an offset of $h = 0.02\ R_S$ (refs 13–15), $r_m = 1.64\ R_S$. This combination of the vertical displacement and the tilt would cause the azimuthal currents to be interrupted at certain positions in the rings for $r > r_m$ (see Fig. 1). (For $r < r_m$ the magnetic latitude of the ring plane is always non-zero and some charged particles will therefore escape absorption by the rings. This disk-population may suffer from pitch-angle scattering through wave–particle interactions and become unstable.) Assuming that the mirror-latitudes of the ions do not change over a time scale of one longitudinal drift around the planet, their drift paths ($r\Delta\Phi$) can be determined by the difference between magnetic latitude (λ_r) of the ring plane and their mirror-latitudes (λ_m). Once the charged particles drift into the region with $\lambda_m > \lambda_r$, they would be absorbed. The variations of $\Delta\Phi$ for two radial distances are sketched in Fig. 2.

Assuming current continuity ($\nabla \cdot j = 0$), the longitudinal variation of the azimuthal current ($j_\perp$) due to the mirroring effect would mean continuous diversion of the currents into the field-aligned components ($j_\parallel$). To estimate the field-aligned current at different longitudes (ϕ) the following relations are needed. First, the current continuity equation can be written as

$$j_\parallel = -\frac{1}{r}\frac{\partial}{\partial\phi}\int j_\perp \, dz \qquad (1)$$

and second, the continuity equation for mirroring particles as

$$\frac{1}{r}\frac{\partial}{\partial\phi}(n_i v_d) = \frac{n_H}{\tau_i} - L_i \qquad (2)$$

with $j_\perp \approx n_i e v_d$ where n_i is the ion number density at longitude ϕ, v_d is the drift velocity[16], n_H is the neutral H atom number density of the ring atmosphere (say), τ_i the ionization time scale and L_i the loss rate. Assuming that the ions can only mirror along the field lines in the region where $\lambda_r > \lambda_m$,

$$n_i \sim n_H\left(\frac{\tau_d}{\tau_i}\right) \qquad (3)$$

where the longitudinal drift time scale between ionization and absorption is $\tau_d = r\Delta\Phi(\phi)/v_d$. Combining these relations, the field-aligned current density at longitude ϕ can be written as:

$$j_\parallel(\phi) \sim er\frac{\partial}{\partial\phi}\int_0^{\lambda_r(\phi)} \frac{n_H}{\tau_i}\Delta\Phi(\phi)\, d\lambda \qquad (4)$$

Note that in the corotating system $\Delta\Phi > 0$ if ions drift westwards towards increasing SLS longitudes (inside the synchronous orbit) and $\Delta\Phi < 0$ if ions drift in the opposite direction (outside the synchronous orbit).

Contour plots of the field-aligned current density ($j_\parallel$) inside the synchronous orbit are shown in Fig. 3. Two maxima are found at SLS longitudes 120° (upward current from ionosphere to the ring) and 250° (downward current from the ring to ionosphere). Such field-aligned currents are corotating with the planet and if the ionization rate in this system is subject to diurnal variation, rotational modulation may conceivably result. For example, the escape flux of the ionospheric electrons reaching the ring at $L = 1.6$ may have large day–night variation due to absence of photoelectron emission and cooling of the exosphere in the night side. (Note that an important nightside heat source, precipitation of energetic particles, does not exist at the latitudes magnetically connected to the rings as the rings are extremely good absorbers of energetic charged particles.)

But note that the azimuthal currents and the associated field-aligned components would be maintained, to a certain degree, even in the shadow of Saturn. This is because the ions produced in the dayside would continue to drift into the shadow to replenish the drift currents. The actual ionization rate must be largely determined by charge exchange between the neutral ring atmosphere and the planetary ionosphere as photoionization and electron impact ionization all have very long ionization time scales ($\gtrsim 10^9$ s). Using the minimum value of $\tau_i \sim 10^6$ s as appropriate for regions a fraction of a planetary radius from the ring plane, and a number density of 500 cm^{-3} for the hydrogen atoms) the peak flux (at SLS longitude 250°) can be estimated as $F_\parallel \sim 10^6\ cm^{-2}\ s^{-1}$ in the equator and $10^7\ cm^{-2}\ s^{-1}$ near the top of the saturnian ionosphere. This is larger than the estimated photoelectron flux of $2 \times 10^5\ cm^{-2}\ s^{-1}$ from Saturn's ionosphere[10]. Thus the downward field-aligned current near SLS longitude 250° could become supercritical, particularly when it moves into Saturn's shadow. In such conditions, an electric field parallel to the magnetic field must be set up accelerating the ionospheric electrons away to maintain the downward current. The parallel potential drop could reach an absolute maximum equal to the total potential drop ψ across the current generating region.[17] This potential ψ arises from the fact that the field-aligned currents must close through the saturnian ionosphere which has a finite conductivity. According to Hill *et al.*[18] the total

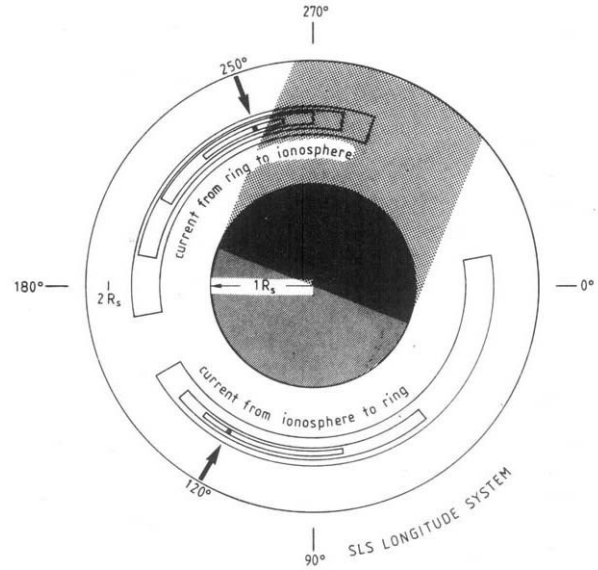


Fig. 3 Contour plot for the parallel electron flux (inside the synchronous orbit) associated with the field-aligned current system. Note that maximum flux from ionosphere to the ring occurs at SLS longitude 250° and the maximum flux from the ring to the ionosphere occurs at 120° SLS longitude.

© 1981 Nature Publishing Group

potential drop across the longitude range $\Delta\Phi$ is

$$\psi = \frac{r\Delta\Phi}{\Sigma_p}\int j_\perp \, dz = \frac{I}{\Sigma_p}\frac{r\Delta\Phi}{\Delta r} \quad (5)$$

Here Δr is the radial range of the field-aligned currents at radial distance r, $\Delta\Phi$ is the corresponding longitude range and Σ_p is the height-integrated Pederson conductivity of the planetary ionosphere. With the total current going through the ring system estimated to be $\sim 10^4$–10^5 A, $\Delta\Phi \sim \pi$, $r \sim 1.5\,R_J$ and $\Delta r \sim 0.3\,R_J$; $\psi = (0.15–1.5)/\Sigma_p$ (MV). Because Σ_p should be of the order of 1 S, if not less, the parallel potentials are conceivably quite large, that is of the order of a few keV.

Although the incorporation of SKR and SED radio emissions into this scheme is tempting but ambiguous a plausible scenario is offered below. We assume that a parallel potential of a few keV could be established in the shadow at SLS longitude 250° and that the ring particles are charged up to electrostatic potential of similar magnitude by the energetic electrons so generated. As soon as the ring particles move into the sunlight there will be immediate electrostatic discharge through emission of photoelectrons (accelerated to keV energies by the negative surface potential)[19]. The electrostatic energy stored on a 1-m radius ring particle charged to 10 keV is 10^{-2} J. Assuming no surface resistance the discharge time would be $\tau \sim 4\times10^{-8}$ s. The total power emitted by each ring particle is 2.5×10^5 W and should be broadband extending to a frequency of roughly $1/\tau \sim 25$ MHz. The maximum intensity of SED should then occur when the spot at SLS longitude 250° emerges from the shadow at which time the SLS longitude 110° faces the Sun (see Fig. 3). This is almost exactly the longitude at which the SED and SKR activities peak. (It is not clear whether the formation of radial spokes in the B ring as observed by the Voyager Imaging experiment[20] is related to this effect.)

SKR events still have to be explained and to a certain extent SKR may also depend on the SED. That is, as the broadband and bursty SED radiation propagates outwards, it will be absorbed by the dense plasma sheet between 4 and 8 R_S (refs 20, 21). If pitch-angle scattering of trapped energetic particles into the atmospheric loss cone is enhanced by plasma wave absorption, rotational modulation of SKR (in the desired manner) would follow. As there are sharp rises in the proton ($E_p <$ 2 MeV) and electron ($E_e < 0.43$ MeV) fluxes near the orbit of Dione[22], precipitation events triggered by the SED radiation must involve principally particles from that region. It is perhaps no great surprise that the SKR emission—besides the 10-h rotational modulation—also seems to be affected by Dione[2].

M.F.T. thanks the Max-Planck-Gesellschaft for hospitality during this work which was in part supported by the NSF under grant ATM 76-82759 and NAS 2-6553.

Received 13 May; accepted 7 July 1981.

1. Warwick, J. W. *et al. Science* **212**, 239 (1981).
2. Gurnett, D. A., Kurth, W. S. & Scarf, F. L. *Science* **212**, 235 (1981).
3. Kliore, A. J. *et al. Science* **207**, 446 (1980).
4. Broadfoot, A. L. *et al. Science* **211**, 206 (1981).
5. Weiser, H., Vik, R. C. & Moos, H. W. *Science* **197**, 755 (1977).
6. Barker, E. *et al. Astrophys. J.* **242**, 383 (1980).
7. Judge, D. L., Wu, F. M. & Carlson, R. W. *Science* **207**, 431 (1979).
8. Carlson, R. W. *Nature* **283**, 461 (1980).
9. Dennefeld, M. *IAU Symp.* No. 65, 471 (1974).
10. Ip, W.-H. *Astr. Astrophys.* **70**, 435 (1978).
11. Ip, W.-H. & Axford, W. I. *Nature* **283**, 180 (1980).
12. Goertz, C. K. *J. geophys. Res.* **85**, 2949 (1980).
13. Smith, E. J. *et al. J. geophys. Res.* **85**, 5651 (1980).
14. Acuña, M. H., Ness, N. F. & Connerney, J. E. P. *J. geophys. Res.* **85**, 5675 (1980).
15. Ness, N. F. *et al. Science* **212**, 211 (1981).
16. Siscoe, G. L. *J. geophys. Res.* **82**, 1641 (1977).
17. Smith, R. A. & Goertz, C. K. *J. geophys. Res.* **83**, 2617 (1978).
18. Hill, T. W., Dessler, A. J. & Maher, L. J. *J. geophys. Res.* (in the press).
19. Mendis, D. A. & Axford, W. I. *A. Rev. Earth planet. Sci.* **2**, 419 (1974).
20. Frank, L. A. *et al. J. geophys. Res.* **85**, 5699 (1980).
21. Bridge, H. S. *et al. Science* **211**, 217 (1981).
22. Trainer, J. H., McDonald, F. B. & Schardt, A. W. *Science* **207**, 421 (1980).

Saturn's kilometric radiation: satellite modulation

M. D. Desch & M. L. Kaiser

NASA/Goddard Space Flight Center, Laboratory for Extraterrestrial Physics, Planetary Magnetospheres Branch, Greenbelt, Maryland 20770, USA

There is an episodic 66-h modulation of the Saturn kilometric radiation which is both frequency and Dione-phase dependent. The behaviour is significantly different from the way in which Io modulates the jovian emission.

THE discovery[1] of nonthermal (kilometric) radio emission (SKR) from Saturn by the Voyager planetary radio astronomy (PRA) instrument has raised the possibility of emission modulation by one or more of its satellites because Saturn, like Jupiter, contains satellites deep within a corotating magnetic field. However, examination of data extending from January to September 1980, yielded no Saturn analogue of the well-known Io-control phenomenon[2]. Instead, short-term (10 day) modulation of the SKR has been reported[3,4] at a period near 66 h, thus implicating satellite Dione as a possible agent. The nature of this modulation, if real, differs dramatically from the way in which Io controls the jovian emission. Using the PRA observations of Saturn, we here search for further evidence of episodic satellite control of SKR and examine the time evolution, frequency dependence and phase of the 66-h modulation. Results will be compared with the phenomenology of the jovian control by Io.

The earlier reports of modulation by Dione presented data in intensity/time format at Dione's revolution period. We will use the method of power spectral analysis to cast the results in a more objective light, although composite intensity–time dynamic spectra will also be used to illustrate graphically the material. The power spectrum method used is that of Deeming[5], and has been applied to the PRA data as described elsewhere[2].

Results

Figure 1 shows the results of power spectral analysis of SKR observed at 174 kHz between 28 October and 18 December 1980. This period of time brackets the Saturn encounter of 12–13 November. The spectrum spans the range of periods from just under 10 h to just above 100 h, corresponding to the range of expected periods of revolution of Saturn's five major innermost satellites and the planetary rotation itself. Aside from the dominant peak near 11 h, which corresponds to modulation of the SKR at Saturn's rotation period, no other statistically significant peaks are evident in the spectra, and, in particular,

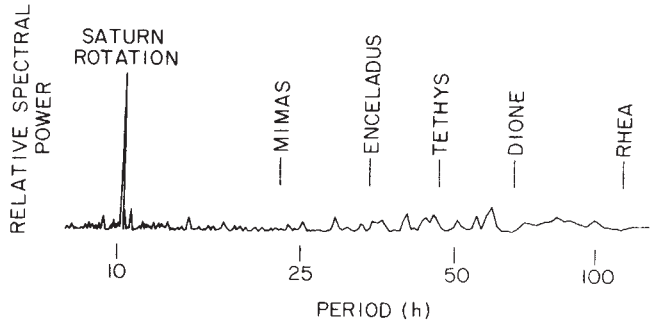


Fig. 1 The results of power spectral analysis of SKR observed at 174 kHz between 28 October and 18 December 1980. Only SKR events exceeding a flux density threshold of 2×10^{-21} W m^{-2} Hz^{-1} (normalized to a standard observer–Saturn distance of 1 AU) were used to eliminate a severe inverse-R-squared bias during this period. The only statistically significant modulation occurs at the 10.66-h Saturn rotation period.

© 1981 Nature Publishing Group

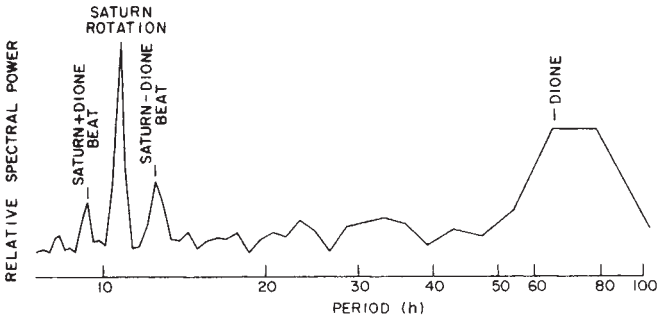


Fig. 2 Power spectral analysis like that described in Fig. 1 but for a shorter interval of time (10–18 November) at 59 kHz observing frequency. Statistically significant peaks occur at the Saturn rotation period, at 66 h, and at the two principal heterodynes between these periods.

there is no significant spectral power at any of the satellite periods. Several other radio frequencies in the SKR band, both above and below 174 kHz, show identical results.

The 66-h modulation

If we examine the PRA low frequencies for the time interval nearer encounter, significant modulation appears near a period of 66 h. To within the 10-h resolution limit inherent in the spectrum near the low-frequency end, this period corresponds to Dione's period of revolution of 65.68 h. In particular, Fig. 2 shows the power spectrum for the analysis interval 10–18 November at an observing frequency of 59 kHz. This 9-day period is equal to three Dione revolutions. The four principal peaks in the spectrum are Saturn's rotation period[2] (10.66 h), Dione (66 h), and the two principal heterodynes between the 10.66 and 66-h periods. A composite dynamic spectrum of the same time interval (Fig. 3*a*) shows the phase of this 66-h modulation. Here, it is clear that the predominant effect is an apparent quenching of the low-frequency SKR near 270° heliocentric phase, that is when Dione is above Saturn's dusk meridian. There is a pronounced gap in the emission occurrence levels at this phase, especially at frequencies below ~200 kHz. Above 200 kHz the emission has begun to fill into the gap, suggesting a frequency dependence of the effect. In addition, because this analysis interval spans the time of closest approach to Saturn, so that the emission is observed from two very different Sun–Saturn–spacecraft angles, the modulation would appear to be locked with respect to the Sun rather than the observer. Note also that there is an additional modulation apparent in Fig. 3. Peaks in the occurrence probability contours occur at 40°, 100°, 160°, 220°, and 340° Dione phase (filled triangles) and are due to the planetary-rotation modulation which is within 3% of an integral multiple of Dione's period. The phase of the expected, but missing, episodes of rotation-modulated SKR are also indicated (open triangle). During this 9-day period, three such episodes fail to occur. It is this gap which produces the peak in the power spectrum at 66 h in Fig. 2.

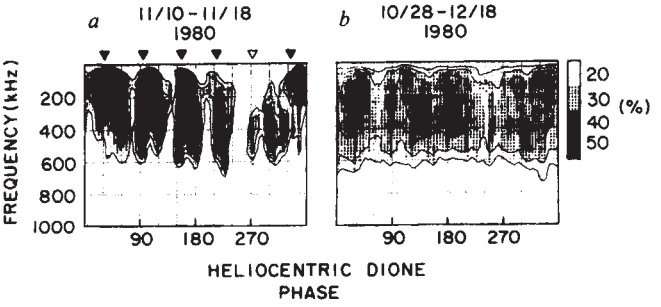


Fig. 3 Composite frequency–time dynamic spectra formed by stacking SKR events at Dione's heliocentric period of revolution for the time interval *a* 10–18 November 1980 and *b* 28 October–18 December. Grey-shaded contour intervals show the resultant per cent occurrence probability levels. The dynamic spectra cover the frequency range 1.2–1,000 kHz. ▼ in *a* show the phase of expected peaks in occurrence levels due to Saturn's rotation; △ shows the 'missing' events. Only events exceeding 2×10^{-21} W m^{-2} Hz^{-1} at 1 AU were used.

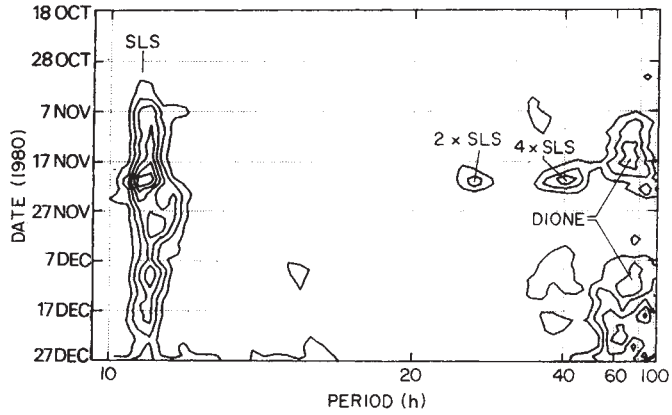


Fig. 4 Results of power spectral analysis on 12-day spans of data from 12 October to 31 December 1980. Spectra were stepped in 1-day increments. Observing frequency was 78 kHz. Contours show levels of relative spectral power as a function of spectral period and centre date of 12-day span. The contour intervals are 15, 25, 35, 45, 55 and 65% of the maximum spectral power. The modulation at Saturn's rotation period (indicated here as SLS, or Saturn longitude system) and two and four times the period are indicated. The broad spectral peak near Dione's period of revolution can be seen to occur in two episodes.

Figure 3*b* supports the long-term power spectral result of Fig. 1. It shows again that when data are examined over extended periods, in this case 28 October to 18 December including the 9-day interval in Fig. 3*a*, the emission occurrence probability levels are distributed randomly in Dione phase. Modulating effects near Dione's period disappear over the long term.

Satellite modulation survey

Before exploring the 66-h modulation further, it is useful to investigate the possibility of additional episodic, short-term modulations. To do this we computed power spectra over 12-day intervals between 12 October and 31 December 1980. Successive spectra were stepped along in 1-day increments, thus yielding an independent spectral estimate on every twelfth point. Each spectrum covered the 90-h window in period spacing from 10 to 100 h (as in Fig. 1). The result is shown in Fig. 4. Contour intervals are proportional to spectral power. The dominant feature is, as might be expected, the Saturn rotation modulation at 10.66 h. It is statistically dominant over almost the entire analysis interval. Outside of this modulation, no significant power appears at periods ≲50 h, except for subharmonic power at 2 and 4 times 10.66 h which appear for a short interval of time over the computation window centred on 21 November 1980 when the modulation at the fundamental is strongest. Above 50 h significant modulation appears twice briefly near 66 h. A third island of significant spectral power slightly below 66 h is centred on 22 December; however, closer examination of the data shows it to be due to an accidental

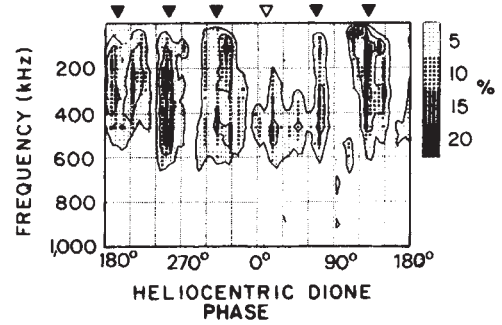


Fig. 5 As in Fig. 3 but for the period 2–14 December 1980. Only events exceeding a flux threshold of 8×10^{-21} W m^{-2} Hz^{-1} (at 1 AU) were used. The 'Dione gap' has shifted to about 30° phase. Note that the abscissa is shifted 180° for clarity.

© 1981 Nature Publishing Group

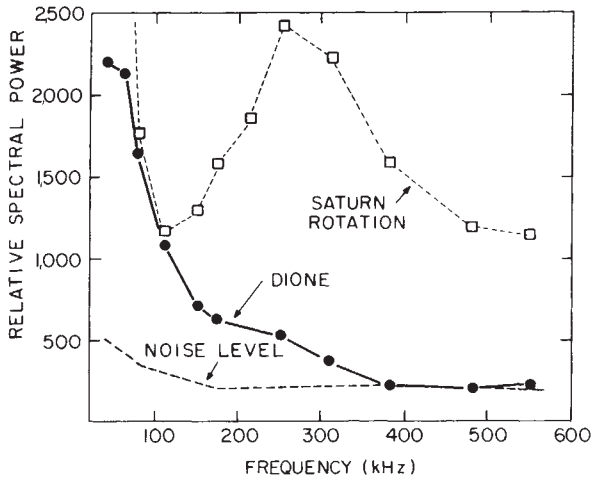


Fig. 6 The decline with increasing frequency in effective modulation at Dione's period of revolution. Spectral power at Dione's period falls to the spectrum noise level at 385 kHz. Modulation due to Saturn's rotation remains strong over the entire band. Analysis interval from 10–18 November 1981. Only events exceeding 2×10^{-21} W m^{-2} Hz^{-1} at 1 AU were included.

reoccurrence of data gaps at this period. At still longer spectral periods additional minor peaks appear; however, their significance is in doubt because they are so near the low-frequency limit of the power spectrum. Identical surveys at several other radio frequencies in the SKR band have yielded identical results. Thus no evidence of short-term modulations of the SKR by any other satellites exists, at least during the late October to December period.

The two significant peaks in Fig. 4 labelled Dione correspond, first, to an interval of time around encounter, mentioned in the discussion of Figs 2 and 3, and, second, to an interval of time in early December. The activity for the period 2–14 December is shown in Fig. 5 in composite frequency–time dynamic spectrum format. The occurrence probability levels are again shown using gray-shaded contour intervals, and the data are organized in heliocentric Dione phase. The expected Dione phase of modulations due to planetary rotation are shown at 10.66-h intervals. The open triangle indicates the 'missing' rotation, in this case 5 missing rotations over the 13-day interval covered. Instead of occurring near 270° Dione phase as was the case near encounter (Fig. 3), the emission gap is now centred on about 30° Dione phase (note shifted phase scale for clarity). The gap is much more pronounced at the lowest frequencies, below 200 kHz, and has disappeared at 400 kHz and higher. This same frequency behaviour was noted with regard to the 66-h modulation observed near encounter.

Frequency dependence

Using power spectra computed at several observing frequencies, we have examined the dependence of the Dione modulation on frequency throughout most of the SKR frequency band. Both Figs 3 and 5 suggested a frequency dependence may exist. The analysis interval extends over the 9-day encounter period shown in Fig. 3 and the results are shown in Fig. 6. Here the power at 66 h is compared with the 10.66-h modulation of the SKR due to Saturn's rotation. The noise level inherent in the power spectra at each frequency is also shown. Note that the magnitude of the 66-h modulation falls by almost a factor of four between 40 and 200 kHz, and reaches statistically insignificant levels at frequencies exceeding ~385 kHz. In total this represents an order of magnitude drop in spectral power over a decade in frequency, f, thus corresponding to a $1/f$ decline. By comparison, the planetary rotation modulation is strong throughout the frequency band, although it also varies with frequency. The rotation modulation approximately follows the shape of the emission flux density spectrum which also has a peak near 200–300 kHz (ref. 6). Below 100 kHz, however, the Saturn rotation modulation increases sharply due, we think, to a possible low-frequency ($f<100$ kHz) emission component[4].

Only those radio events which exceeded a constant flux density threshold, in this case 2×10^{-21} W m^{-2} Hz^{-1} normalized to an equivalent observer–Saturn distance of 1 AU, were used to derive the frequency dependence shown in Fig. 6. Because there is some dependence of spectral power on the choice of flux threshold used, we examined the effect on the frequency-dependence curve of varying this threshold. We found that, over a two decade range of flux densities, the modulation is always greatest at low frequencies with no statistically significant modulation appearing at $f>400$ kHz. The results shown for early December (Fig. 5) are clearly consistent with this frequency dependence. We conclude that the 66-h modulation is strongly and inversely dependent on frequency.

Discussion

In the two intervals of time during which significant modulation occurs near a spectral period of 66 h the modulation (1) varies inversely with frequency and (2) persists for short intervals limited to 3–5 cycles of the 66-h period. The phase of the modulation is not constant, however. The observed phase of the emission gap changes from about 270° Dione phase for the November episode to ~30° phase for the December episode.

Phenomenologically this modulation differs dramatically from the way in which Io is observed to control the jovian emission (DAM). Io enhances rather than quenches the emission level of DAM; the Io effect is, to first order, independent of frequency; the phase of the emission is fixed with respect to the observer, not the Sun; and, finally, Io control is not an intermittent phenomenon.

If Dione is responsible for the 66-h modulation of the SKR, it is difficult to imagine what mechanism could cause such an effect. Certain features, namely the cutoff in observed emission level and the $1/f$ frequency dependence, are consistent with preferential absorption or strong refraction of SKR at low frequencies due to a moderately dense plasma associated with Dione. In fact, Pioneer 11 (ref. 7) and Voyager[3,8] observations seem to support the existence of a 15 to 20 cm^{-3} electron plasma in the vicinity of 4–8 R_S. Dione, at 6.3 R_S, is within this plasma. However, other observed features of the modulation are inconsistent with this simple picture. One might expect, for example, a more pronounced modulation on the inbound observing period when Voyager, at a saturnigraphic latitude of +8°, viewed the emission from a point closer to the equator plane. Instead, the outbound geometry seems more favourable, when Voyager was at +26° latitude. Also, during the first episode of the Dione modulations, one of the SKR gaps occurred near closest approach when Voyager 1 was at high southern latitudes. Furthermore, the variable phase of the modulation makes any simple absorption picture difficult to model.

An alternative method of cutting off the emission includes a more dynamic interaction between Dione and the radio source region. There is some evidence that Dione may be stimulating field-aligned currents, perhaps as Io does at Jupiter[9]. If so, Dione may act, in some as yet unspecified way, to quench the emission process physically. The Dione flux tubes do intersect the modelled[6] radio source region; however, the orbital phase of the intersection is near 120° Dione phase, not the preferred 270° or 30° gap locations. Thus neither the absorption nor dynamic quenching models are especially attractive, and we hope that the Voyager 2 observations of SKR near closest approach will shed some light on this perplexing problem.

Received 13 May; accepted 7 July 1981.

1. Kaiser, M. L. *et al. Science* **209**, 1238–1240 (1980).
2. Desch, M. D. & Kaiser, M. L. *Geophys. Res. Lett.* **8**, 253–256 (1981).
3. Gurnett, D. A., Kurth, W. S. & Scarf, F. L. *Science* **212**, 235–239 (1981).
4. Warwick, J. W. *et al. Science* **212**, 239–243 (1981).
5. Deeming, T. J. *Astrophys. Space Sci.* **36**, 137–158 (1975).
6. Kaiser, M. L., Desch, M. D. & Lecacheux, A. *Nature* **292**, 731–733 (1981).
7. Frank, L. A. *et al. J. geophys. Res.* **85**, 5695–5708 (1980).
8. Sittler, E. C., Scudder, J. D. & Bridge, H. *Nature* **292**, 711–714 (1981).
9. Acuña, M. A., Connerney, J. E. C. & Ness, N. F. *Nature* **292**, 721–724 (1981).

© 1981 Nature Publishing Group

Control of Saturn's kilometric radiation by Dione

W. S. Kurth*, D. A. Gurnett*, & F. L. Scarf†

*Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA
†TRW Defense and Space Systems, One Space Park, Redondo Beach, California 90278, USA

Voyager 1 observations of Saturn's kilometric radio emissions reveal a strong but apparently transitory control by the orbital phase angle of Dione. This may be a geometrical effect and a time-variable plasma torus associated with Dione could explain most of the observed details of the Dione modulation by creating a shadow zone near the equatorial plane.

THE Voyager planetary radio astronomy investigation provided the first incontrovertible evidence of nonthermal radio emissions from Saturn[1]. The radiation spectrum peaks in the kilometre wavelength regime near 200 kHz and periodicities in the occurrence of the radio bursts give an internal planetary rotation period of 10 h, 39.9 min (ref. 2). Observations of the radio emission at 56.2 kHz by the plasma wave instrument on Voyager 1 revealed a modulation of the intensity of the radio emission with a period very close to the orbital period of Dione indicating that some control of the emission was a function of the orbital phase of the moon[3]. This control was also reported by the planetary radio astronomy investigation[4] with the maximum effect seen at lower frequencies[5], but extending as high as 385 kHz. The magnitude of the Dione influence is not constant in time and was strongest near closest approach. (Dione has a radius of 560 km (ref. 6) and is 6.29 R_S from Saturn in an orbit whose eccentricity is 0.002.)

Here we analyse the effect of Dione's orbital phase on the emission of radio waves from Saturn to explain the apparent transitory nature and suggest a mechanism for the interaction with the moon.

Nature of Dione's influence

We begin by describing the saturnian kilometric radio (SKR) spectrum below ~56 kHz. Figure 1 is a typical event as seen by the plasma wave instrument as Voyager approached Saturn on 11 November 1980. The amplitude of radio signals as a function of time for the upper five channels of the plasma wave spectrum analyser[7] are shown. The height of the solid black area represents the average power flux (W m^{-2} Hz^{-1}).

The radio burst shown in Fig. 1 increases in amplitude with increasing frequency to a peak at least as high as the upper frequency limit of the instrument (56.2 kHz). The emission is highly time-variable, changing in amplitude by nearly an order of magnitude on time scales of a few minutes. While gross amplitude changes track fairly well from one channel to the next, many of the fine-scale features are totally uncorrelated with those of adjacent channels indicating narrowband elements in the spectrum. Hence, the temporal variations of the SKR are similar in many respects to the auroral kilometric radiation from the Earth. The low-frequency cutoff of the saturnian emission has been reported[1] to be frequently near 60 kHz; however, Fig. 1 shows that the emission is detectable at frequencies as low as 10 kHz.

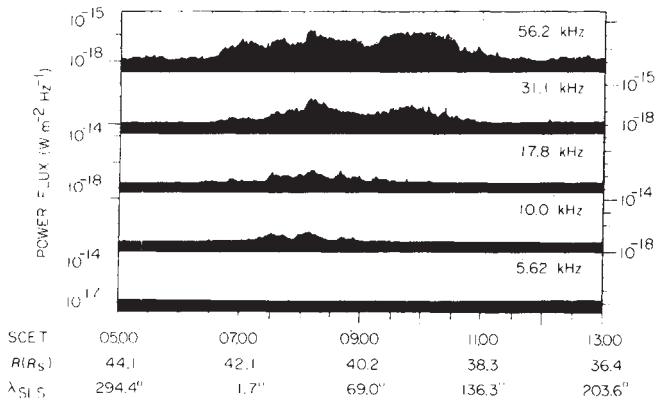


Fig. 1 A typical example of saturnian kilometric radiation extending from near 10 kHz to above 56 kHz.

We specifically consider the correlation between the orbital phase angle of Dione, ϕ_{Dione}, and the amplitude of the kilometric radio emission which was originally demonstrated by Gurnett *et al.*[3], who showed that in addition to the 10 h 39.9 min modulation related to Saturn's rotation, the emission was strongest when Dione was in the local dawn sector. Figure 2 plots the hourly average power flux of the radio emission at 56.2 kHz as a function of radial distance; also shown is a dashed line with a R^{-2} dependence. The arrows indicate times when Dione was at local dusk ($\phi_{Dione} = 270°$) and correspond to periods when there is little or no kilometric emission. Figure 2 suggests that the effect of Dione is to suppress or attenuate the saturnian radio emission when $\phi_{Dione} \simeq 270°$. (The nearly-constant minimum signal strength of $\sim 8 \times 10^{-19}$ W m^{-2} Hz^{-1} corresponds to the receiver noise level.) The peaks of the 10 h 39.9 min bursts approximately follow a R^{-2} dependence except for periods within ~200 R_S when the amplitude of the peaks fall well below the curve coincident with the passage of Dione through the dusk sector. Beyond ~200 R_S on the inbound leg and ~165 R_S on the outbound leg the periodic suppression in phase with Dione's orbital motion is no longer evident. The Dione suppression might explain the apparent north–south asymmetry in the source strength at lower frequencies reported by Warwick *et al.*[4] since Dione was near local dusk for most of Voyager 1's trajectory through southern latitudes.

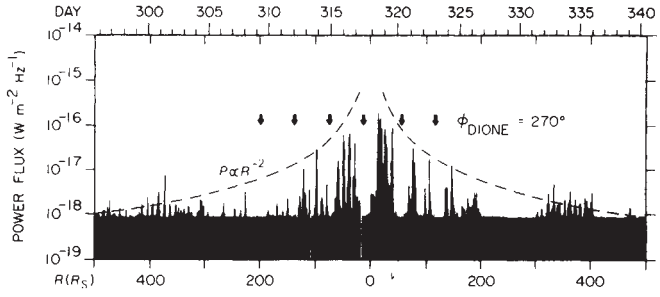


Fig. 2 A plot of hourly average values of the 56.2 kHz plasma wave receiver channel as a function of radial distance from Saturn. The narrow bursts are emitted at a period of ~10 h 39.9 min and generally rise in amplitude with decreasing R approximately as R^{-2}. The arrows indicate times when Dione is at local dusk and correspond closely to periods when there is little or no radio emission.

Analysis of traces similar to that in Fig. 7 of ref. 3, and the present Fig. 2, shows that a clear signature of the Dione modulation is not present before day 308 or after day 325 of 1980. These dates correspond to Voyager–Saturn distances of 227 and 165 R_S, respectively. To illustrate the transitory nature of the Dione effect further, we show the result of a statistical study of the occurrence of kilometric radiation as a function of Dione's

0028–0836/81/340742—04$01.00

© 1981 Macmillan Journals Ltd

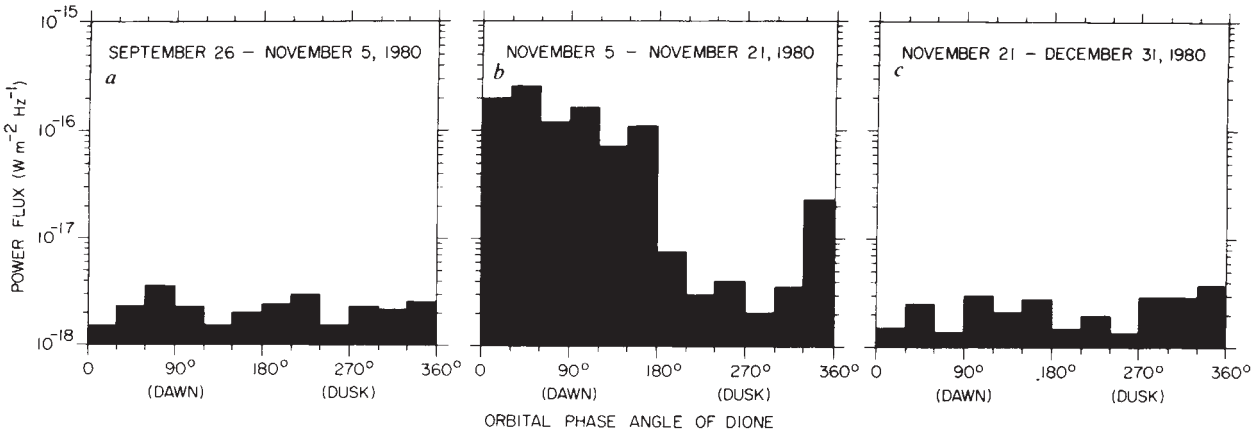


Fig. 3 Plots of the 98% amplitude as a function of Dione's orbital phase angle for the distant inbound, near encounter, and distant outbound Voyager 1 trajectories at 56.2 kHz. Note the strong Dione control near encounter and the complete lack thereof during the distant inbound and outbound passes.

orbital phase angle. The bar graphs in Fig. 3 show the power flux below which 98% of the samples fell for each of 12 phase angle bins. Recall that phase angle is measured positive eastward from a plane containing Saturn's rotation axis and the Sun, with 0° at local midnight. Figure 3*b* includes all data covered by Gurnett *et al.*[3] when the Dione effect is qualitatively most apparent. Figure 3*a*, *c* present results from the inbound and outbound legs when little or no Dione control was apparent. It is clear the distant inbound and outbound data show no significant trend. On the other hand, the difference in amplitude at the 98% level between phase angles near 90° and 270° is about two orders of magnitude for the two-week period near closest approach. For this period, Dione clearly has an important role in determining the amplitude of Saturn radio emissions at 56.2 kHz.

Discussion

There are many possible interactions by which a moon may influence the generation of magnetospheric radio emissions. We shall attempt to narrow the list of likely candidates to one or two which are consistent with all or most of the observations. We assume the SKR originates at relatively low altitudes on high latitude field lines, presumably in the auroral region as suggested by Kaiser *et al.*[8].

We first need to explain the transitory nature of the effect. A truly transient process which is active for two weeks and then inactive for long periods of time requires a mechanism to stimulate the control. The long-term studies show no Dione effect[4] so the active periods must be quite rare. Hence, it is strange that the effect was seen coincidentally with closest approach and we shall argue that the effect is not transitory but depends on the location of the observer.

The simplest explanation of the transitory nature of the Dione control is geometric in nature and specifically related to Voyager's low latitude or small distance from the equatorial plane of Saturn around closest approach. The effect became apparent around day 308 when Voyager 1 was at 8.3° latitude and about 33 R_S above the equatorial plane. On day 325, when the effect apparently disappeared, the spacecraft was at a latitude of 25.4° and 71 R_S above the equatorial plane. A second episode of Dione control in early December has been reported by the planetary radio astronomy team[5]. The effect is visible in the plasma wave data at 56.2 kHz, however, we are not confident in interpreting our data as a genuine Dione effect at this time. Should this be a valid episode, however, our conclusion of a geometric effect would be in jeopardy.

Figure 2 showed Dione attenuated the saturnian radio emission as opposed to stimulating or amplifying it. The attenuation could be caused by a propagation effect or by a basic change in the source itself. In either case, the most likely interaction mechanism is a change in magnetospheric density associated with the satellite. A decrease in density could only be accomplished by a sweep-up effect and this is hardly a time-variable process capable of disturbing the inner magnetosphere on time scales of a few days. An increase in magnetospheric density, however, is plausible as the result of a process which liberates particles from the moon making the moon a plasma source. This type of process occurs in the jovian system with Io being the primary source of plasma.

Frank *et al.*[9] argued that Dione could be a reasonable source for an oxygen torus at Saturn. A peak in the ion density at Dione's orbit with density and temperature profiles reminiscent of those at Io in Jupiter's magnetosphere suggests that Dione, like Io, is a plasma source. Also, the water ice or frost surface of Dione opens the possibility of plasma production through dissociation and ionization. A ledge-like structure in the plasma torus near Dione's L-shell has been reported[10], and similarities of this structure with that observed at Io in Jupiter's magnetosphere suggest a plasma source. Evidence from the Pioneer 11 magnetometer[11] also suggests a process involving plasma production. On the inbound Pioneer trajectory, significant magnetic field perturbation was seen at Dione's L-shell but no effect was observed on the outbound leg. Local time of Dione was ~ 21 h ($\phi_{\mathrm{Dione}} \sim 315°$) and ~ 1 h ($\phi_{\mathrm{Dione}} \sim 15°$) for the Pioneer inbound and outbound Dione L-shell crossing, respectively. Although there are several possible explanations, this magnetic field signature is consistent with a ring current associated with Dione which varies with ϕ_{Dione}.

An increase in magnetospheric density could affect radio emissions by modifying conditions at the source of the radiation. There is some evidence that the Earth's auroral kilometric radiation is quenched by increasing the density in the source region[12–14], however, Gurnett and Anderson[15] have stated that this is not very likely. Another difficulty with direct modification of the source is that the plasma produced at Dione would have to have easy access to the source. If the radio source were in the auroral region, it is unlikely that an equatorial plasma source at $L \sim 6$ could effectively modify the density in the auroral zone.

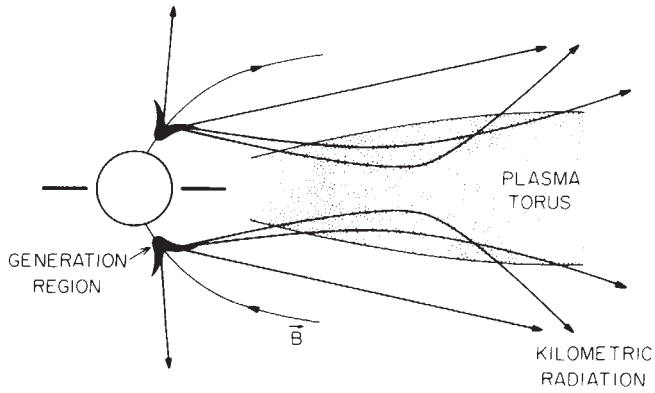


Fig. 4 How a Dione-related torus would refract radio waves from Saturn away from the equatorial region. A similar effect was found for broadband kilometric radiation from Jupiter.

© 1981 Nature Publishing Group

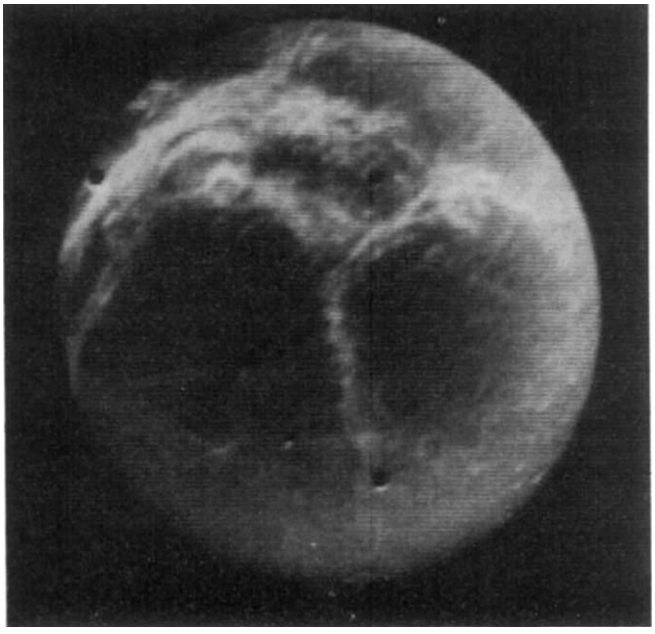


Fig. 5 An image of the trailing hemisphere of Dione taken by Voyager 1 when Dione was near local dusk. Note that this sunlit hemisphere is dominated by bright wispy features which are probably fresh outcroppings of water ice or frost. The frost may be a source of plasma for a torus whose density fluctuates with the same period as Dione's orbit.

The alternative is to assume that an increase in plasma density associated with Dione affects the propagation of radio waves through refraction or reflection as represented in Fig. 4. We used the model torus from ref. 3, although the model shown by Bridge *et al.*[10] would give essentially the same effect. Waves with frequency less than or equal to the local plasma frequency in the torus cannot penetrate the torus. Waves of higher frequency will be refracted away from the equatorial region, although the effect will diminish with increasing frequency.

Many observations suggest that the model shown in Fig. 4 is viable. First, the density of the plasma torus as measured by Pioneer 11 (ref. 9) and Voyager 1 (refs 3, 10) is of the order of 40 cm^{-3} corresponding to a plasma frequency of 57 kHz. Waves below this frequency could not penetrate the torus and waves well above 100 kHz could undergo substantial refraction. This refraction is consistent with the frequency dependence of the Dione effect reported by Desch and Kaiser[5], as lower frequencies tend to be refracted most. Second, because preliminary observations[3,10] suggest the plasma is confined near the equator, waves propagating near low latitudes would be most affected. Therefore, we expect a low-latitude shadow zone as was found for the jovian kilometric radiation[16,17]. A shadow zone near the equator would explain the presence of a Dione control only near encounter when Voyager 1 was close to the equator. The asymmetry in the latitude at which the effect appears and disappears could well be a local time asymmetry in the thickness or density of the torus.

How then does the orbital phase angle of Dione modulate the emission? It is important first to establish that the modulating effect of the torus is longitudinally symmetric. The evidence is embodied in Fig. 7 of ref. 3 and Fig. 2 of this paper. Note that the phase of the Dione effect did not change as Voyager moved from near local noon before encounter to local early morning after the encounter. If there were a localized cloud of plasma which orbited Saturn, a phase shift would have been apparent as the viewing point changed. As the phase did not change, the modulation effect must be longitudinally symmetric. This longitudinal symmetry should be fairly easily achieved due to the rapid corotation of the plasma which will quickly distribute plasma in longitude even though the source may be localized.

To explain the dependence of the torus density on the orbital phase angle of Dione, we suggest that Dione may be a plasma source only when it is near local dusk. Hence, the plasma cloud produced dissipates on a time scale of about a day so that by the time Dione is near local dawn the torus density has decreased by a factor of say, two. Because the trailing hemisphere of Dione is always the same, a hemispheric asymmetry in direct particle sputtering processes cannot explain the dawn–dusk asymmetry. On the other hand, there are obvious asymmetries in the surface features of Dione. Figure 5 shows Dione near local dusk. The trailing hemisphere shows a complex pattern of wispy features which do not appear on the opposite hemisphere of Dione. Smith *et al.*[6] suggest the bright markings are controlled by a regional system of fractures or faults possibly formed or reopened by internally generated stresses. The bright material is probably water ice and it is tempting to speculate that this relatively fresh surface frost is photosputtered when the trailing hemisphere is sunlit as it is when Dione is near local dusk, or that photodissociated volatiles produce the plasma.

Carlson[18] and Frank *et al.*[9] have discussed the rings as possible sources of plasma in the saturnian magnetosphere. Of processes such as photosputtering, ion sputtering, sublimation with subsequent photodissociation, and others, photosputtering seems to be the most efficient mechanism. It would be a convenient mechanism for an orbital phase-dependent source of plasma at Dione in view of the hemispherical asymmetry of the wispy features shown in Fig. 5. Taking the hydrogen atom flux derived by Carlson[18] of $\sim 10^8\ cm^{-2}\ s^{-1}$, Dione can produce $\sim 10^{24}$ atoms s^{-1} (assuming the entire sunlit hemisphere is a source and photosputtering is applicable to Dione). If the time scales for both creation and dissipation of the plasma torus are ~1 day (to provide a 2.74-day modulation period), then the production/loss rate must be the order of $10^{28}\ s^{-1}$. We have assumed a torus 4 R_S thick, centred at 6 R_S, which fluctuates between 20 and 40 cm^{-3}. Obviously, photosputtering is not sufficient to produce a dense torus on the time scale of a day.

If we assume that the other processes considered by Carlson are no more efficient than photosputtering, we must consider a mechanism which is based on the release of volatiles from fractures in Dione's crust and subsequent photodissociation. Presumably the volatiles would be released into the exosphere at a constant rate but preferentially from the wispy hemisphere. Photodissociation of the molecules would occur only during the dusk sector of Dione's orbit when the wisps are sunlit. This mechanism is largely speculative and production rates depend on rate of volatile release, the type of volatile, and photodissociation rate. However, the injection rate at Io has been estimated[19] at $2 \times 10^{29\pm1}$ ions s^{-1}, about 200 times that required at Dione.

The proposed fluctuating torus model also requires a dissipation rate of the order of $10^{28}\ s^{-1}$. Richardson *et al.*[19] point out that outward radial diffusion through flux–tube interchange is the predominant loss mechanism at Jupiter and given that oxygen or some other heavy ion is an important constituent of the Dione torus[9,10] the centrifugally-driven interchange instability is likely to be important at Saturn, also.

We have arrived at a model which is similar in part to effects observed at Jupiter. We propose that Dione is responsible for the production of plasma when it is near local dusk which forms a longitudinally symmetric torus. The torus then casts a radio shadow on low latitudes so that the saturnian radio emission cannot be seen by a spacecraft near the equator. The torus decays with a time constant of about one day so that as Dione approaches local dawn the usual saturnian emission reappears. The possibility of an additional Dione-control episode[5] in early December, 1980 at $\phi_{Dione} \sim 40°$, however, presents an even more perplexing situation to be explained involving a shift in the phase of the control.

The research at The University of Iowa was supported by NASA through contract 954013 with the Jet Propulsion Laboratory and through grants NGL-16-001-002 and NGL-16-001-043 with NASA Headquarters, and by the Office of Naval Research. The research at TRW was supported by NASA through contract 954012 with the Jet Propulsion Laboratory.

© 1981 Nature Publishing Group

Received 13 May; accepted 17 July 1981.

1. Kaiser, M. L., Desch, M. D., Warwick, J. W. & Pearce, J. B. *Science* **209**, 1238–1240 (1980).
2. Desch, M. D. & Kaiser, M. L. *Geophys. Res. Lett.* **8**, 253–256 (1981).
3. Gurnett, D. A., Kurth, W. S. & Scarf, F. L. *Science* **212**, 235–239 (1981).
4. Warwick, J. W. *et al. Science* **212**, 239–243 (1981).
5. Desch, M. D. & Kaiser, M. L. *Nature* **292**, 739–741 (1981).
6. Smith, B. A. *et al. Science* **212**, 163–191 (1981).
7. Scarf, F. L. & Gurnett, D. A. *Space Sci. Rev.* **21**, 289–308 (1977).
8. Kaiser, M. L., Desch, M. D. & Lecacheux, A. *Nature* **292**, 731–733 (1981).
9. Frank, L. A., Burek, B. G., Ackerson, K. L., Wolfe, J. H. & Mihalov, J. D. *J. geophys. Res.* **85**, 5697–5708 (1980).
10. Bridge, H. S. *et al. Science* **212**, 217–224 (1981).
11. Rairden, R. L. thesis, Univ. Iowa, (1981).
12. Benson, R. F. & Calvert, W. *Geophys. Res. Lett.* **6**, 479–482 (1979).
13. Wu, C. S. & Lee, L. C. *Astrophys. J.* **230**, 621–626 (1979).
14. Calvert, W. *Geophys. Res. Lett.* (in the press).
15. Gurnett, D. A. & Anderson, R. R. in *Physics of Auroral Arc Formation* (in the press).
16. Kurth, W. S., Gurnett, D. A. & Scarf, F. L. *Geophys. Res. Lett.* **7**, 61–64 (1980).
17. Green, J. L. & Gurnett, D. A. *Geophys. Res. Lett.* **7**, 65–68 (1980).
18. Carlson, R. W. *Nature* **283**, 461 (1980).
19. Richardson, J. D., Siscoe, G. L., Bagenal, F. & Sullivan, J. D. *Geophys. Res. Lett.* **7**, 37–40 (1980).

Pre-encounter distributions of Saturn's low frequency radio emission

T. D. Carr*, J. J. Schauble† & C. C. Schauble†

* Department of Astronomy, University of Florida, Gainesville, Florida 32611, USA
† University of Florida Graduate Center, Eglin Air Force Base, Florida 32542, USA

An analysis of Voyager 1 pre-encounter data is presented in which one-month averages of flux density from Saturn are determined as functions of both central meridian longitude and frequency. Comparisons of corresponding distributions for two one-month intervals seven months apart yield information on their stability, and a redetermination of the magnetospheric rotation period.

THE detection of a nonthermal component of Saturn's thermal radio emission has proved to be difficult. No synchrotron radiation from the planet has yet been found, and pre-Voyager searches for a nonthermal component at low frequencies were inconclusive. An unsubstantiated identification of such emission[1] was made near 1 MHz, but not until Voyager 1 had approached to within 3 AU was a nonthermal radio component positively identified[2]. This radiation is not unlike the kilometric bursts from the terrestrial and jovian magnetospheres. Most of the activity is between ~70 and 300 kHz, and is polarized predominantly in the right hand sense[2–4]. The probability of occurrence tends to vary in a more or less repeatable way with

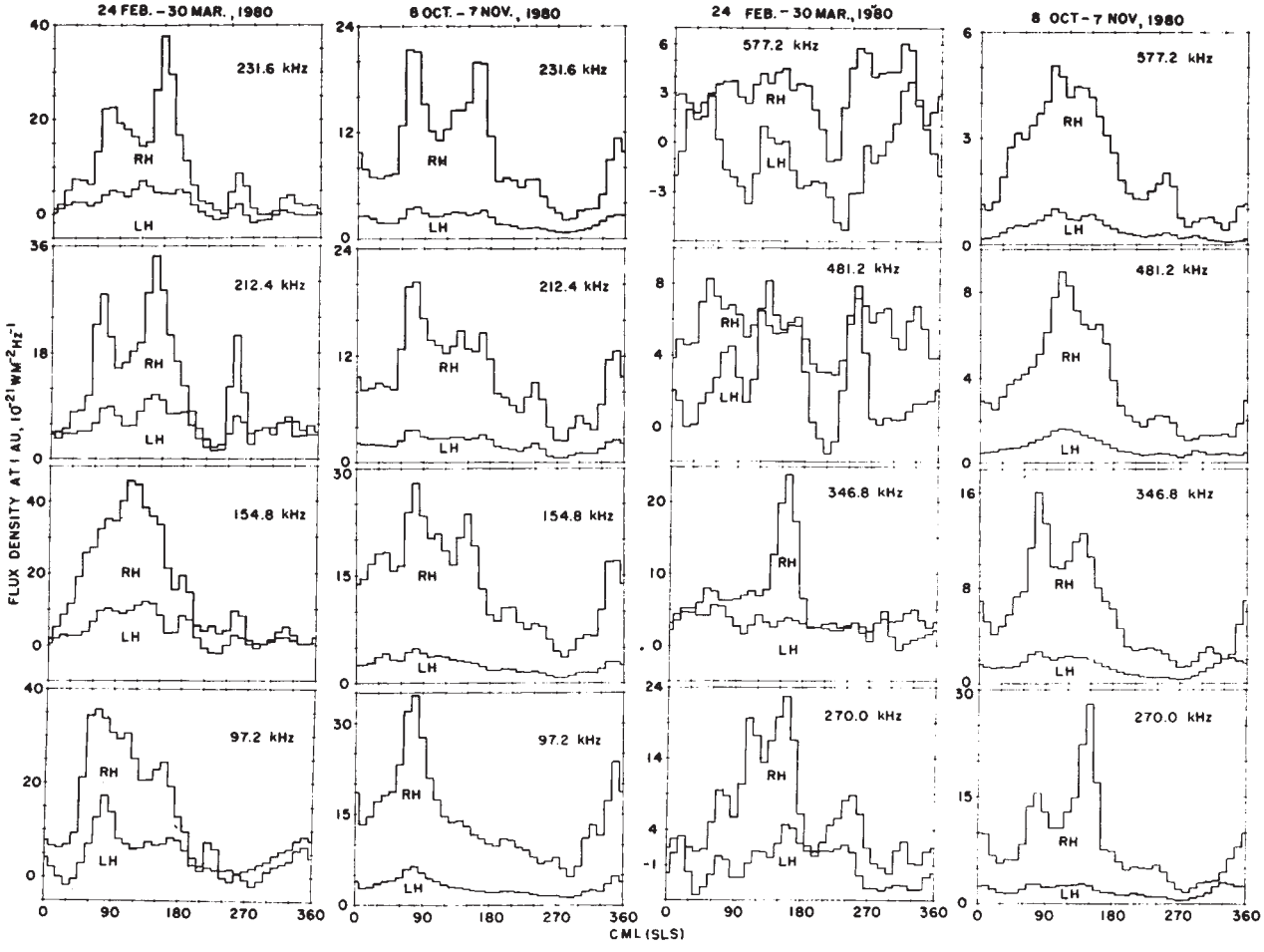


Fig. 1 Histograms of average flux density per 10° interval as a function of central meridian longitude for specified time intervals and frequencies. Averaging extends over inactive as well as active periods. Right hand (RH) and left hand (LH) polarized components are plotted separately. Small and negative flux densities result from background fluctuations when Saturn radiation is very weak, accentuated at times by slight errors in baseline determination.

0028–0836/81/340745—03$01.00

© 1981 Macmillan Journals Ltd

the rotation of the planet. From an analysis of this periodicity over several months before the Voyager 1 encounter, Desch and Kaiser[3] arrived at a rotation period of 10 h 39 min 24(±7) s, the presently accepted value for Saturn's magnetosphere. This period is the basis for the Saturn longitude system (designated SLS) in current use for specifying central meridian longitude (CML).

Observations

The data used in this analysis were obtained by Voyager 1 during two periods extending from 24 February to 30 March 1980, and from 8 October to 7 November 1980. Closest approach to Saturn was on 12 November 1980. The measurements were made with the low band receiver of the planetary radio astronomy experiment[5], at 28 selected frequencies from 20.4 to 1,038 kHz. In calculating flux density as functions of CL and frequency, the averaging intervals covered inactive as well as active periods. Data reading and manipulation was computerized, except for preliminary editing when data containing interference from within the spacecraft, from the Sun, or from Jupiter were excluded. Solar bursts were easily identified but Jupiter bursts, which more closely resemble those from Saturn, were identified in most cases by locating the same event on the Voyager 2 record and noting the difference of onset times. As Voyager 2 was much closer to Jupiter than Voyager 1, nearly every jovian event appearing on a Voyager 1 record could also be seen on the corresponding Voyager 2 record with greater intensity and earlier onset time. For each of the two circularly polarized components at each frequency one or more short segments of undisturbed baseline were selected from each 24-h run (except for a very few runs in which no activity lull occurred). Zero-flux density baselines fitted to these segments were subtracted from the data points within that 24-h interval. Following the background subtraction and the application of an appropriate calibration factor, flux densities were normalized to a standard 1 AU spacecraft distance. Histograms of average flux density for 10° CML intervals formed the basis for subsequent analyses.

Results and discussion

Eight pairs of representative histograms are shown in Fig. 1. The two in each pair are for the same frequency, one for the first and the other for the second data period. Although the CML regions of maximum and minimum activity approximately coincide for histograms at the same frequency, the differences seen are probably due to statistical fluctuations. An extreme case of this is seen in the 481.2 and 577.2 kHz histograms, in which the radiation during the first data period is so weak that background fluctuations dominate. At the same two frequencies for the second data period, when the spacecraft was close to the planet, the distribution is very clearly defined.

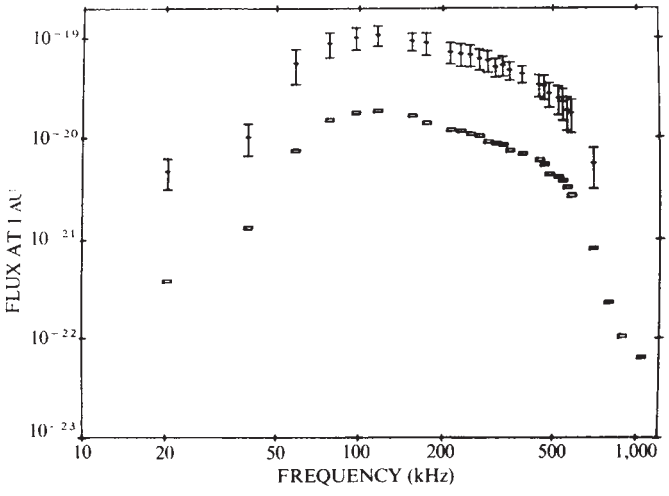


Fig. 2 Special distributions of average flux density (lower curve) and of 'peak' or 97 percentile flux density (upper curve). Error bars indicate 90% confidence range for the latter. Flux densities in $Wm^{-2} Hz^{-1}$.

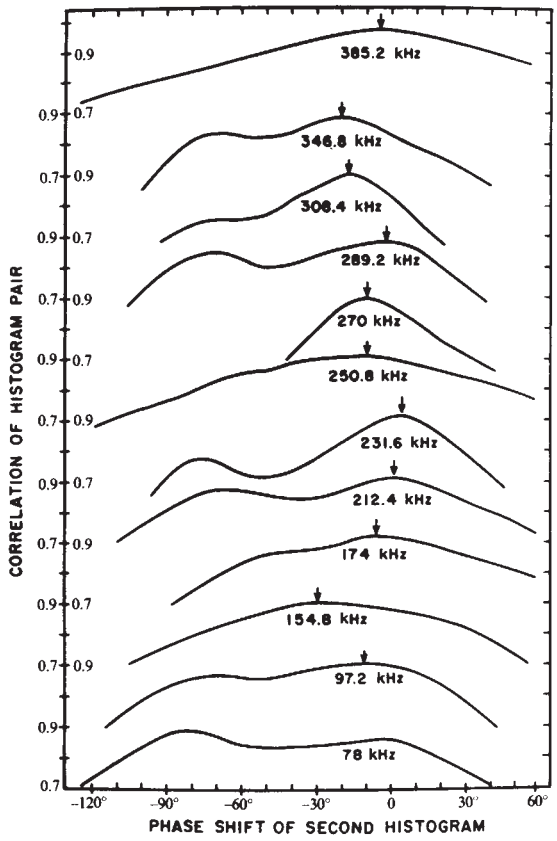


Fig. 3 Cross correlation of pairs of single-frequency histograms as a function of the longitude shift of the second histogram in the pair. Curve maxima are indicated by arrows.

In addition to statistical fluctuations, there seem to be systematic differences. For example, a secondary peak often appears near 360° in the second histogram which is not present in the first one. Such differences are unlikely to be latitude effects. The saturnigraphic latitude of the spacecraft on the median dates of the two data periods was 9.1° and 8.7°, respectively, unlikely to cause an observable effect: nor was the difference in saturnian local time at the spacecraft enough. The systematic differences are most likely due to temporal instability possibly caused by some alteration of the magnetospheric plasma distribution, and even of the multipole structure of the magnetic field, over the seven-month interval between the two data sets.

Figure 1 shows that the right hand (RH) circularly polarized component is much stronger than the left hand (LH) one. The histograms suggest that the ratio of LH to RH flux densities may be smaller in the vicinity of the main peak (near 100°) than elsewhere; however, this ratio would be subject to large errors

Table 1 Rotation period measurements

Frequency (kHz)	Shift (deg)	Mean	Correlation coefficient	Mean	Rotation period (10 h 39 min+)
97.2	−11		0.90		
154.8	−30		0.90		
174.0	−5		0.92		
		−15		0.91	21 s
212.4	1		0.91		
231.6	3		0.90		
250.8	−10		0.90		
270.0	−10		0.90		
		−4		0.90	23 s
289.2	−2		0.87		
308.4	−17		0.88		
346.8	−21		0.87		
385.2	−5		0.96		
		−11		0.89	22 s

Weighted mean rotation period: 10 h 39 min 22 s. Maximum deviation of the three values from the mean: 1 s. SLS rotation period (Desch and Kaiser[3]): 10 h 39 min 24(±7) s.

© 1981 Nature Publishing Group

outside the main peak region because of the relatively low flux density values there.

The spectral distribution of average flux density (RH and LH combined), shown in Fig. 2, was obtained by integrating the histograms of average flux density against CML for each of the 28 frequencies for the second data set. This curve closely agrees with that corresponding to the first data set, except at those frequencies at which the signal-to-noise ratio in the first set was poor. Nearly all of the radiation occurs below 1 MHz, and the broad spectral peak has its maximum at ~150 kHz. Figure 2 also shows the spectral distribution of the 97-percentile level of the individual flux density measurements (RH and LH combined) in the second data set. This is the level which is exceeded 3% of the time. Each point on the 97-percentile curve is the mean of the 24-h 97-percentile values for the 1-month period.

Integration of the average curve of Fig. 2 yields the 1 AU equivalent of the average power per unit area (5.3×10^{-15} Wm^{-2}) reaching the spacecraft at all frequencies. The corresponding isotropic equivalent average power of the saturnian radio source is 1.5×10^{9} W. This is ~0.4% of the isotropic power equivalent for all the jovian emission below 40 MHz, or about five times that of the jovian kilometric emission (below 300 kHz)[6]. A similar integration of the 97-percentile curve in Fig. 2 could not be used to calculate a 97-percentile isotropic equivalent power, because emission peaks do not occur simultaneously at all frequencies.

Single-frequency pairs of histograms such as those in Fig. 1 allow a more precise determination of Saturn's magnetospheric rotation period. Our method was previously used successfully to redetermine Jupiter's radio rotation period[7]. Each histogram in a single-frequency pair represents data from a period of about a month, and the interval between the medium dates of the two data periods was ~224 days (1.935×10^{7} s). The histograms are based on the SLS rotation period[3]. From the measured shift of the second histogram in each pair with respect to the first, the correction to be applied to the SLS period was calculated. The histogram for the RH component was used in each case although the sum of the RH and LH components could have been used instead. The correct histogram shift was indicated by the principal maximum of the cross correlation function for the pair. The corrected rotation period, P_{new}, was then calculated from the formula

$$P_{new}=P_{SLS}\left(1+\frac{P_{SLS}\Delta L}{360T}\right)$$

where P_{SLS} is the SLS period, ΔL is the longitude shift (in degrees) of the second histogram with respect to the first at the cross correlation maximum, and T is the time between the median dates of the two data sets from which the histograms were plotted. The signal-to-noise ratio during the first data period was considered adequate for cross-correlation of 12 of the 28 originally selected frequencies. Plots of the regions about the principal peaks of the cross-correlation functions for the histogram pairs at these 12 frequencies are shown in Fig. 3. The principal peaks in 11 curves occur between −30° and 4°. Most of the curves also exhibit a secondary peak in the vicinity of −80°. For 78 kHz the positions of the principal and secondary peaks are interchanged; this curve was rejected. The longitude shifts at the cross-correlation maxima for the 11 remaining curves, and the correlation coefficients are presented in Table 1. These 11 measurements are not completely independent, because whenever activity is observed on a given frequency channel, it most likely occurs at about the same time on one or more nearby channels. Desch and Kaiser (personal communication) have found the $1/e$ decorrelation bandwidth of the Voyager pre-encounter Saturn emission to be ~115 kHz. We therefore averaged our 11 shift measurements in three groups, the centre frequencies of which are separated by nearly this amount, as indicated in Table 1. The resulting averages, which lie within ~1 s of the mean (10 h 39 min 22 s), are believed to be largely independent. The mean differs from the SLS period by only 2 s. The close agreement of our three measurements suggests that their statistical precision is better than the ±7 s quoted for the SLS determination. However, Desch and Kaiser believe that this uncertainty is due to a basic instability in the averaged pre-encounter saturnian emission pattern. If this is true it suggests that the close agreement of our measurements is accidental. Further rotation period determinations over different time intervals should settle this question.

To approach maximum statistical precision using our method, a near-optimum combination should be sought for three intervals: (1) the data collection interval for the first histogram; (2) that for the second histogram (assumed to be as close to encounter as possible without producing excessive latitude change), and (3) the interval between the median dates of the two histograms. The latter should provide an adequate temporal baseline over which to measure the rotation period, yet not make the Saturn events in the first interval too weak or too infrequent. We believe that a large part of the difference in the shapes of our first and second histograms in each pair results from the relatively low signal intensity relative to the background noise during the first interval. We should be able to improve our precision (and evaluate it) soon by using more recent Voyager 1 and 2 data.

We thank M. D. Desch and M. L. Kaiser for unpublished decorrelation bandwidth data, and for helpful comments.

Received 13 May; accepted 8 July 1981.

1. Brown, L. W. *Astrophys. J.* **198**, L89–L92 (1975).
2. Kaiser, M. L., Desch, M. D., Warwick, J. W. & Pearce, J. B. *Science* **209**, 1238–1240 (1980).
3. Desch, M. D. & Kaiser, M. L. *Geophys. Res. Lett.* **8**, 253–256 (1981).
4. Warwick, J. W. *et al. Science* **212**, 239–243 (1981).
5. Lang, G. J. & Peltzer, R. G. *IEEE Trans* AES-**13**, 466–471 (1977).
6. Alexander, J. K., Carr, T. D., Thieman, J. R., Schauble, J. J. & Riddle, A. C. *J. geophys. Res.* **86** (in the press).
7. May, J., Carr, T. D. & Desch, M. D. *Icarus* **40**, 87–93 (1979).

Plasma wave turbulence at planetary bow shocks

F. L. Scarf*, D. A. Gurnett† & W. S. Kurth†

* Space Science Department, TRW Defense and Space Systems Group, Redondo Beach, California 90278, USA
† Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA

Voyager 1 observations of plasma wave turbulence at Saturn's bow shock are discussed and compared with corresponding data from Jupiter, Earth, and Venus. The results suggest that the plasma instabilities that develop at the lower Mach number bow shocks of the terrestrial planets differ from those found at the high Mach number bow shocks of the outer planets.

AS the solar wind streams out from the Sun, its flow speed remains constant, but its plasma density, ion and electron temperatures, and mean interplanetary magnetic field strengths all have distinct and varying radial profiles. Thus, the dimensionless interplanetary parameters that determine how microscopic plasma processes develop change with heliocentric distance. We therefore expect to find significant radial variations in several important interplanetary plasma phenomena. This is very important to collisionless shock studies, in particular, because some combinations of plasma parameters occurring naturally in the solar wind have not yet been attained experimentally.

The planetary bow shock also forms in the solar wind, and this means that the characteristics of these very high Mach number discontinuities will vary across the Solar System as the expanding wind encounters each planet at a different radial position. As

0028-0836/81/340747—04$01.00

© 1981 Macmillan Journals Ltd

the processes that develop at the bow shock directly affect the properties of the post-shock plasma that actually impacts the ionopause or magnetopause, comparative planetary studies must include comparison of the bow shock turbulence spectra and associated wave–particle interaction phenomena. We now discuss recent Voyager 1 observations of plasma wave turbulence at Saturn's bow shock, and present an initial comparison of these wave measurements with corresponding data from Jupiter, Earth and Venus. The results suggest that the plasma instabilities that develop at the lower Mach number bow shocks of the terrestrial planets differ from those found at the high Mach number bow shocks of the outer planets. Thus the state of the post-shock plasma varies significantly from one planet to another.

Saturn bow shock crossings

As noted by Gurnett *et al.*[1], the solar wind upstream from Saturn was remarkably quiet in the 10 Hz–56 kHz range covered by the Voyager 1 plasma wave instrument. The only regular signals observed in this region were repetitive series of Saturn radio emissions[1,2]. We also detected a single cluster of intense 3 kHz noise bursts between 17.05 and 17.20 on 11 November 1980, when the Voyager to Saturn distance was 32.4 R_S. We interpret these 3-kHz waves as electron plasma oscillations associated with suprathermal electrons streaming back from the shock, but when these emissions disappeared, Voyager was still in the solar wind. Terrestrial experience with similar sequences of observations suggests that the bow shock was actually near the spacecraft just after 17.00, but that the shock surface receded back towards the planet in response to a change in solar wind conditions. Similar 3-kHz bursts were detected again after 23.22, and these waves persisted until 23.27 when Voyager made its only inbound bow shock crossing at a radial distance of 26.1 R_S. Figure 1 shows all of the 16-channel wave measurements for this event, and it can be seen that the shock was thin and well-defined in terms of the low frequency ($f \leq 1$ kHz) wave activity. These bandpass channel amplitude profiles are very similar to those found when Voyager 1 approached Jupiter[3,4], and in both cases, the post-shock magnetosheath noise levels seem to be exceptionally low.

Just before 12.27, the wind speed (V) and density (N) were[5] 420 km s^{-1} and 0.11 ions cm^{-3}, and the magnetic field strength (B) was[6] about 0.3 nT. Here, the upstream proton and electron temperatures were $\sim 1.5 \times 10^4$ K and 2.3×10^4 K, respectively (J. Scudder, personal communication). For these parameters, the Alfvén speed, $V_A (=[B^2/4\pi N m_+]^{1/2})$, is only 20 km s^{-1}, and the magnetosonic wave speed, V_{ms} $(=[V_A^2 + V_s^2]^{1/2}$, with $V_s = [(5kT_+/3 + kT_-)/m_+]^{1/2})$, is ~ 28 km s^{-1}. Hence, the Mach numbers for the collisionless shock of Fig. 1 are remarkably high ($M_{ms} = V/V_{ms} = 15$, and $M_A = V/V_A = 21$), and this event is probably one of the highest Mach number collisionless shocks for which any *in situ* observations are available.

Before comparing plasma wave measurements for the relatively well-defined dayside bow shock crossings at Saturn, Jupiter, Earth and Venus, we discuss characteristics of the outbound bow shock encounter with Saturn on the nightside

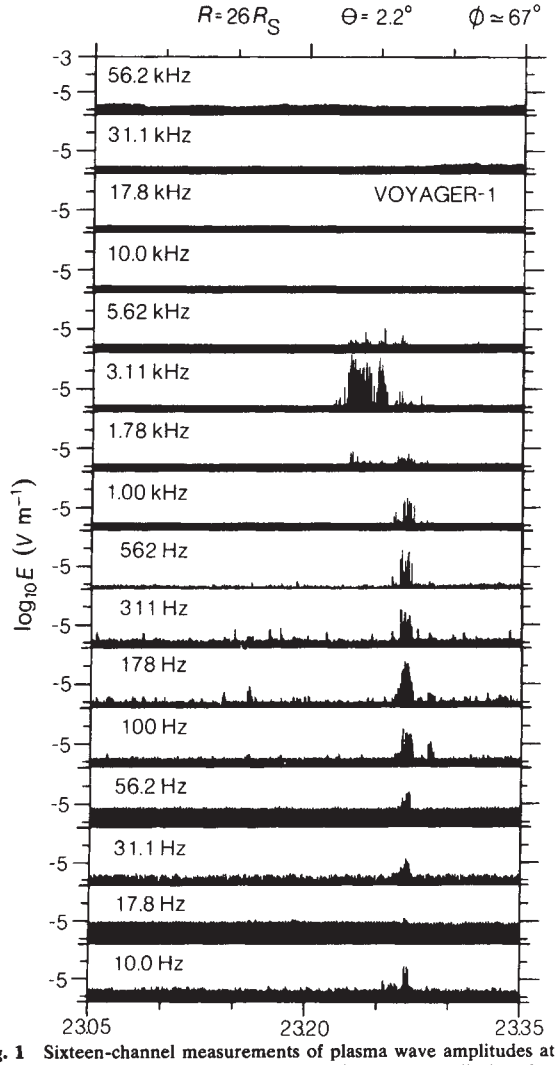


Fig. 1 Sixteen-channel measurements of plasma wave amplitudes at the inbound crossing of Saturn's bow shock. Several sporadic interference effects have been deleted. The 31.1-kHz and 56.2-kHz levels represent radio emissions from Saturn, and we interpret the 3.1-kHz waves as electron plasma oscillations.

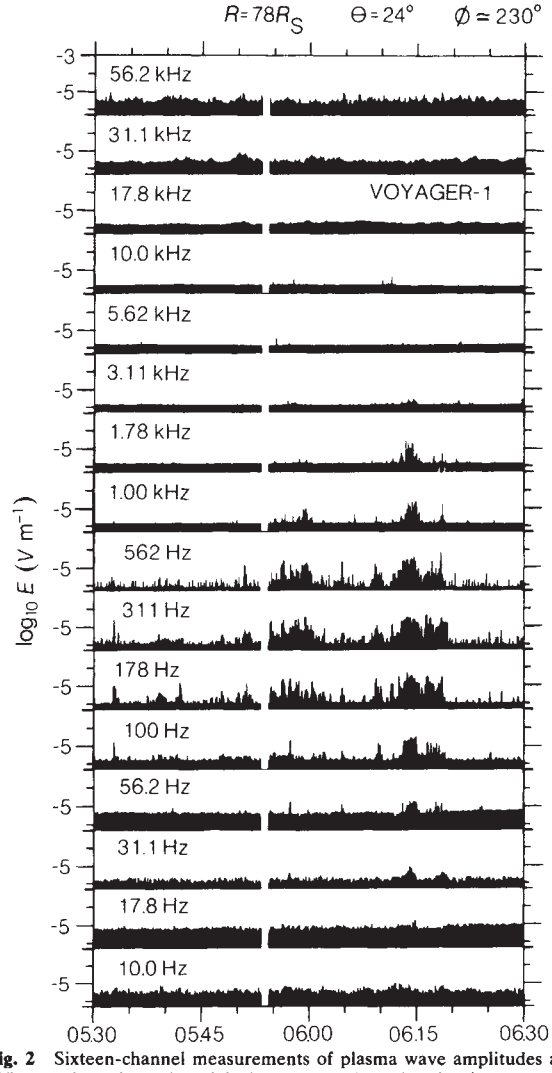


Fig. 2 Sixteen-channel measurements of plasma wave amplitudes at the diffuse outbound crossing of the bow shock. Here, the 10–56 kHz waves are Saturn radio emissions, and the gap between 05.53:23 and 05.54:11 corresponds to an interval when the Voyager tape recorder was acquiring waveform data.

© 1981 Nature Publishing Group

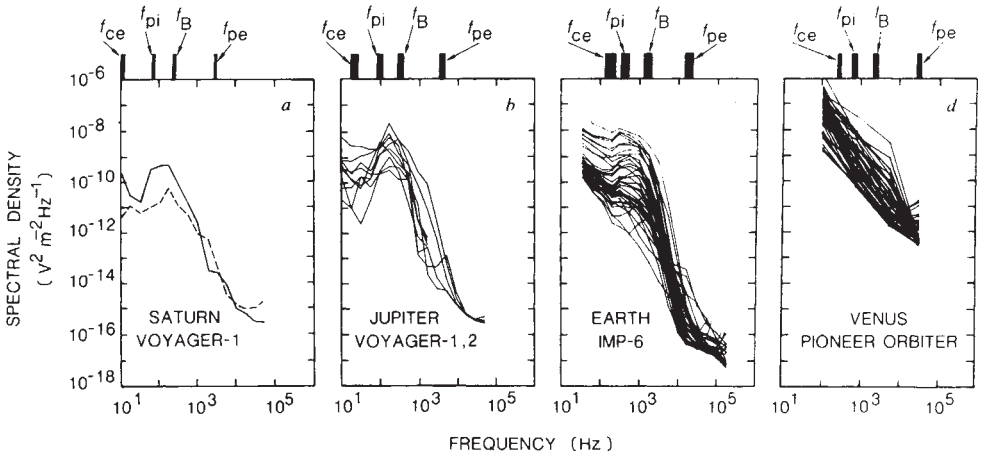


Fig. 3 Multi-planet comparison of plasma wave spectra. *a* Was constructed using the 16-channel amplitude measurements of Figs 1 and 2. The solid and dashed spectra represent the inbound and outbound crossings, respectively. The characteristic frequencies shown at the top are the electron cyclotron frequency (f_{ce}), the ion plasma frequency (f_{pi}), the Buneman mode frequency (f_B), and the electron plasma frequency (f_{pe}). Although the spectra for Earth (*c*), Jupiter (*b*) and Saturn (*a*) all have a primary or secondary peak in the noise spectrum between f_{pi} and f_B, the intense low frequency enhancement which is evident in most terrestrial spectra and all the Venus observations is not found at Jupiter or Saturn.

almost midway between local midnight and dawn. During the interval 05.50 to 06.18 on 16 November when Voyager 1 was at 78 R_S, the spacecraft traversed a relatively diffuse transition layer separating the magnetosheath from the solar wind. Figure 2 shows the relative thickness of the outbound shock and the variability in the plasma wave amplitude profiles; this is characteristic of a quasi-parallel shock configuration that is usually associated with an extensive region of upstream disorder, and it seems that after 06.18, much turbulence was present for an extended period. Ness *et al.*[6] noted the elevated r.m.s. magnetic field noise levels which indicate that enhanced low frequency electromagnetic wave activity was present after 06.18. The plasma wave instrument was able to detect Saturn-associated bursts of ion acoustic waves and electron plasma oscillations out to at least early December more than 400 R_S away from Saturn. These results are readily understandable because at 9.5 AU, the Parker spiral model leads to an average interplanetary field oriented almost perpendicular to the Saturn–Sun line. Thus, the Saturn 'upstream region' or 'foreshock', which can develop when the *B*-field intersects the shock surface, was encountered after closest approach rather than before.

Planetary bow shock turbulence spectra

Pioneer Venus Orbiter provided the first data on plasma wave activity at the bow shock of a planet other than Earth in December 1978, and in March and July 1979, the Voyager spacecraft transmitted corresponding information from Jupiter's bow shock. Following the 1980 Voyager 1 encounter with Saturn, it is possible to present a preliminary comparison of wave–particle interactions at four planetary bow shocks, but a comprehensive analysis must await Voyager 2 Saturn data.

The IMP-6 study of Rodriguez and Gurnett[7] serves as a reference for analysis of plasma waves at the terrestrial bow shock, and Fig. 3*c* shows the average electric field spectra detected at 36 representative crossings of the shock surface. For these crossings, the average solar wind density was 5.2 particles cm^{-3}, and the average magnetic field strength was 7.5 nT; these values yield a mean electron plasma frequency ($f_{pe} = 9{,}000\sqrt{N}$) of 20.5 kHz, an ion plasma frequency ($f_{pi} = 210\sqrt{N}$ for protons) of 480 Hz, and an electron cyclotron frequency ($f_{ce} = 28B$) of 210 Hz. These characteristic frequencies, along with the 1.65 kHz value that is appropriate for the corresponding Buneman mode[8] characteristic frequency, f_B ($=[m_e/m_i]^{1/3} f_{pe}$) are marked at the top of Fig. 3.

Rodriguez and Gurnett studied how the plasma wave spectrum varied with changes in Alfvén and sonic Mach numbers, plasma β-value ($\beta = 8\pi NK[T_e + T_i]/B^2$), ion temperature, electron-to-ion temperature ratio, solar wind density and shock normal angle. Plots of the mid-frequency (200 Hz–4 kHz) *E*-field amplitude against T_e/T_i showed a strong positive correlation, and a weaker one when the low frequency (20 Hz–200 Hz) *E*-field amplitude was used, but the extreme shapes of the shock spectra shown in Fig. 3*c* were not discussed in association with any particular groups of upstream plasma parameters. Indeed, as the upstream parameters associated with the Rodriguez and Gurnett data set varied over a huge range ($0.5 \leq N \leq 19\ cm^{-3}$, $2 \leq B \leq 17$ nT, $1.5 \leq M_A \leq 26.6$, $0.03 \leq \beta \leq 4.5$), there was no *a priori* reason to expect to find characteristic changes in shock spectra as we moved inward towards Venus or out to Jupiter and Saturn. Nevertheless, Fig. 3*a*, *b*, *d* strongly suggests that the bow shock spectra at the other planets differ significantly from those detected at Earth.

The 48 Venus spectra on Fig. 3*d* have been discussed elsewhere[9]: these spectra are limited because the Pioneer Venus instrument has only four filter channels and it uses short antennas which are affected by spacecraft interference. Nevertheless, the shock processes at Venus seem to generate much higher levels of low frequency ($f \leq f_{pi}$) plasma wave turbulence than do the corresponding processes at Earth [we use $N(\text{Venus}) \simeq (12 \pm 4)$ electrons cm^{-3} and $B(\text{Venus}) \simeq (10 \pm)$ nT to compute the characteristic frequencies shown at the top of Fig. 3*d*].

Recently the nine jovian spectra shown in Fig. 3*b* have been compared with the IMP-6 data[10]. The Voyager 1 and 2 measurements were all obtained during the dayside inbound passes, and the characteristic frequencies shown at the top correspond to $N \cong (0.225 \pm 0.075)$ electrons cm^{-3} and $B \simeq (0.8 \pm 0.2)$ nT. The low frequency enhancement in electric field intensity, evident in most of the terrestrial wave spectra and all of the Venus observations, was not present at Jupiter. Assuming that these low frequency waves represent whistlers, it was speculated that whistler-mode noise may not be as important the jovian bow shock as at Earth and Venus. Alternatively, the Jupiter spectral peaks at higher frequencies could be attributable to plasma conditions in the distant solar wind that yield enhanced generation of ion acoustic or Buneman mode oscillations, rather than suppressed levels for the whistlers.

Figure 3*a* shows the Voyager 1 Saturn bow shock spectra derived from the measurements of Figs 1 and 2; the solid lines connect one-minute averages for the inbound crossing, and the dashed lines refer to the outbound case. The Jupiter and Saturn shock spectra are quite similar and both have intensity peaks near the f_{pi} and f_B characteristic frequencies. It is difficult to identify these wave modes, however, as ion acoustic waves have phase speeds small compared with the solar wind speed; thus, Doppler effects could account for large frequency shifts, and these waves may have $f \leq f_{pi}$ in the solar wind rest frame. As the heliocentric distance increases, there appears to be a significant and consistent variation in the average bow shock turbulence, and we should therefore examine how some relevant interplanetary parameters vary from one planetary orbit to another.

A complete investigation of the variations in bow shock microstructure requires comparative study of the electron and ion temperatures, thermal anisotropies, heat flux moments, and distribution function shapes, as well as comparison of the macroscopic upstream parameters such as magnetic field vector, solar wind velocity, and wind density. Note that the average spectra shown in Fig. 3 can be associated with changes in average

© 1981 Nature Publishing Group

Mach number. Specifically, using $M_A = V/V_A$ and $V_A = (B/(4\pi N m_+^{1/2})$, the mean B and N values cited above give $M_A(\text{Venus}) = 6.3$, $M_A(\text{Earth}) = 5.6$, $M_A(\text{Jupiter}) = 11$, and $M_A(\text{Saturn}) = 21$. This suggests that there may be a relation between Mach number and spectral shape.

Support for this speculation comes from reanalysis of the IMP-6 bow shock observations. Although the result is not strongly evident in Fig. 3*c*, three of these spectra have 'outer planet' shapes in the sense that for these cases, the wave level at $f \approx 100$ Hz is much lower than the level at $f \approx 100$–1,000 Hz. These peaked terrestrial spectra were detected when the upstream conditions yielded a subset of high Mach numbers (P. Rodriguez, personal communication), and this tends to support a hypothesis that the Jupiter and Saturn shock spectra have unusual characteristics, primarily because the solar wind Mach numbers at the outer planets tend to be much higher than the customary values near 0.7–1.0 AU.

Discussion

On theoretical and experimental grounds it is known[11,12] that several distinct dissipation mechanisms develop in collisionless shocks and that the dominant interaction depends strongly on the Mach number. For low Mach numbers ($M \leq M^* \approx 2$–3), the shock is resistive, and for $M > M^*$, viscous dissipation becomes important. Above a second critical Mach number, $M^{**} \approx 6$–7, theory[11] requires other types of dissipation, but these very high Mach number shocks have not been intensively analysed. We now find that upstream conditions at the outer planets generally yield Mach numbers considerably higher than M^{**}, and thus it is of great interest to find that the Jupiter and Saturn bow shock crossings have plasma turbulence spectra that resemble those found at Earth when M is high. The Jupiter and Saturn characteristics are also unusual in that the jump in electron temperature across the shock is enormous ($T_e^2/T_e^1 \approx 15$ at Jupiter[13]) and the magnetosheath wave levels are extremely low.

In conclusion, our study of the bow shocks at Jupiter and Saturn has provided insight into collisionless shock microstructure for an important new plasma regime, and helps to understand shocked plasmas of non-equilibrium characteristics that seem to be different from those found in the terrestrial magnetosheath.

We thank J. Scudder, P. Rodriguez, F. Coroniti and C. Kennel for valuable discussions. The research at TRW was supported by NASA under Contract 954012 with Jet Propulsion Laboratory, and the research at The University of Iowa was supported by NASA through contract 954013 with Jet Propulsion Laboratory, through grants NGL-16-001-002 and NGL-16-001-043 with NASA headquarters, and by the Office of Naval Research.

Received 13 May; accepted 7 July 1981.

1. Gurnett, D. A., Kurth, W. S. & Scarf. F. L. *Science* **212**, 235 (1981).
2. Warwick, J. W. *et al. Science* **212**, 239 (1981).
3. Scarf, F. L., Gurnett, D. A. & Kurth, W. S. *Science* **204**, 991 (1979).
4. Scarf, F. L., Gurnett, D. A., Kurth, W. S. & Poynter, R. L. *Nature* **280**, 796 (1979).
5. Bridge, H. S. *et al. Science* **212**, 217 (1981).
6. Ness, N. F. *et al. Science* **212**, 211 (1981).
7. Rodriguez, P. & Gurnett, D. A. *J. geophys. Res.* **81**, 2871 (1976).
8. Buneman, O. *Phys. Rev. Lett.* **1**, 8 (1958).
9. Scarf, F. L., Taylor, W. W. L., Russell, C. T. & Elphic, R. C. *J. geophys. Res.* **85**, 7599 (1980). (1980).
10. Gurnett, D. A. & Scarf, F. L. in *Physics of the Jovian Magnetosphere* (ed. Dessler, A. J.) (Cambridge University Press, in the press)
11. Biskamp, D. *Nucl. Fus.* **13**, 719 (1973).
12. Formisano, V. *J. Phys.* **38**, C6 (1977).
13. Scudder, J. D., Sittler, E. C. Jr & Bridge, H. S. *J. geophys. Res.* **86** (in the press).

Surface waves on Saturn's magnetopause

R. P. Lepping*, L. F. Burlaga* & L. W. Klein†

* NASA/Goddard Space Flight Center, Laboratory for Extraterrestrial Physics, Greenbelt, Maryland 20771, USA
† Computer Sciences Corporation, 8728 Colesville Road, Silver Spring, Maryland 20910, USA

Voyager 1 magnetometer data have shown that small-amplitude surface waves occurred on Saturn's dayside magnetopause, causing multiple inbound crossings of this boundary. These waves were travelling approximately parallel to Saturn's equatorial plane along the magnetopause ('tailward'), suggesting that they were driven by the rotation of Saturn's magnetosphere. Hydromagnetic waves (possibly slow mode) were observed in the adjacent magnetosheath.

ON 11 November (day 316) 1980 at 23.27 UT Voyager 1 crossed Saturn's bow shock on its inbound trajectory towards the planet. This was followed by a series of five, fairly regularly spaced magnetopause (MP) crossings ending at ≈02.47 UT of day 317, and discussed in a preliminary report[1]. The analysis and interpretation of the magnetometer data taken over this ≈3.5-h period is now discussed in terms of the first observations of surface waves on Saturn's magnetopause, their characteristics, and a plausible connection between the magnetopause and hydromagnetic waves occurring in the magnetosheath.

Observations

The dual three-axis low field magnetometer system onboard Voyager 1 was in its most sensitive range of 0.0043 nT per quantization step for the entire period of interest. The measurements, at 0.060-s intervals, were probably accurate to ±0.03 nT per axis. The magnetometer experiment is described in detail elsewhere[2]. Figure 1 presents 9.6-s averages of the magnetic field and shows the inbound bow shock, five magnetopause crossings, and the intervening magnetosheath. The inset shows the trajectory of the spacecraft projected on the orbital plane of Saturn; in the inbound magnetosheath the spacecraft was only about 10° below this plane (in a Saturn centred system). The two outstanding characteristics in the data are: (1) large quasi-periodic oscillations in magnitude (B) of the field in the magnetosheath that seem to grow in amplitude from the bow shock to the MP; (2) fairly regularly spaced MP crossings.

Table 1 lists the centre times and durations (Δt) for the five MP crossings. The field latitudes are $\langle\delta\rangle \approx -75°$, on the magnetosphere side of each crossing, as expected for Saturn's distant front-side magnetosphere[1,3], and the field shows moderate draping at medium (positive) latitudes in the magnetosheath near the MP. That is, the magnetosheath field near the MP boundary is apparently being constrained to lie parallel to the MP as it is convected along the boundary. At more distant locations from the MP the magnetosheath field is closer to $\delta \approx 0°$ at $\langle\lambda\rangle \approx 285°$. The directional changes of the field across the MPs were large, permitting an accurate estimation of the normal to the MP surface for each crossing.

Table 1 Inbound magnetopause crossings and normals

No.	day 317 centre time (h min s)	Δt (min)	Method 1 δ_N	Method 1 $\lambda_N - 180°$	Method 1 η	Method 2 δ_n	Method 2 $\lambda_n - 180°$	Method 2′ δ_{eq}	Method 2′ $\lambda_{eq} - 180°$
1	01.54:45	2.50	−14°	17°	82°	−4°	2°	−3°	4°
2	02.14:23	2.25	17°	−16°	78°	7°	−3°	5°	−6°
3	02.28:43	0.92	−11°	48°	89°	−11°	47°	8°	47°
4	02.42:28	1.08	30°	−20°	78°	23°	−4°	19°	−14°
5	02.46:51	0.43	−23°	38°	89°	−22°	40°	−4°	45°
		Average	0°	13°	83°	−1°	16°	5°	15°
		r.m.s.	23°	31°	6°	17°	25°	9°	29°

0028–0836/81/340750—04$01.00

© 1981 Macmillan Journals Ltd

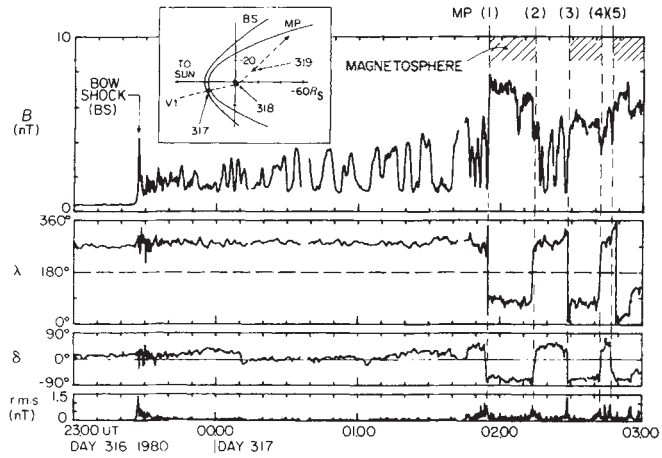


Fig. 1 Voyager 1 magnetic field (9.6 s) averages for 4 h around Saturn's inbound magnetosheath in terms of field magnitude B, longitude λ, latitude δ, and pythagorean r.m.s. deviation. The angles λ, δ are expressed in a heliographic, spacecraft centred, coordinate system, such that $\lambda = \tan^{-1}(B_T/B_R)$ and $\delta = \sin^{-1}(B_N/B)$, where: $\hat{\mathbf{R}}$ is radially away from the Sun; $\hat{\mathbf{T}}$ is parallel to the Sun's equator plane, normal to $\hat{\mathbf{R}}$, and positive in the general direction of Saturn's orbital motion; and $\hat{\mathbf{N}} = \hat{\mathbf{R}} \times \hat{\mathbf{T}}$. At this time $\hat{\mathbf{N}}$ is parallel to the ecliptic plane normal within ~2°. The five inbound magnetopause crossings are labelled. The day numbers in the trajectory insert refer to spacecraft positions at the start of the referred day, and $R_S \equiv 60{,}330$ km, Saturn's radius.

Analysis of the magnetopause

Two methods were used to estimate normals to the MP for comparison purposes. The first[4] uses a variance analysis of difference fields ($\Delta\mathbf{B}_i$) across the discontinuity and assigns the normal direction (δ_N, λ_N) to the minimum variance direction; $\Delta\mathbf{B}_i = (\mathbf{B}_i - \langle\mathbf{B}\rangle)$, $\mathbf{B}_i$ is one of the series of 1.92-s averages (in our case) occurring in the analysis interval across the MP, and $\langle\mathbf{B}\rangle$ is the average of the $\mathbf{B}_i$s. The second method[5] is similar to the first but uses the full vectors, $\mathbf{B}_i$ (1.92-s average fields); then the minimum variance direction of the $\mathbf{B}_i$s is the estimated normal direction (δ_n, λ_n) (see Table 1). The angle η is defined as $\cos^{-1}(B_z/\langle B\rangle)$, where B_z is the component of $\langle\mathbf{B}\rangle$ along the estimated normal, according to method (1), and $\langle B\rangle$ is $|\langle\mathbf{B}\rangle|$. When $\eta = 90°$ the MP discontinuity is described as a tangential discontinuity. Within the expected errors on the ηs all crossings can be considered as tangential discontinuities; this is also common for the Earth's dayside MP. When the MP is a tangential discontinuity and when magnetometer zero level errors can be ignored, method (2) yields more accurate normal estimates than method (1) (ref. 6). Only the results of method (2) will be used in our surface wave analysis. However, both methods qualitatively show the same main features: oscillations in both δ and λ, and almost the same overall averages.

Both methods indicate that surface waves were occurring on the MP surface, just as have been seen to occur on the Earth's MP[7,8]. The nature of these waves is more evident in a more appropriate coordinate system. The variance analysis of method (2) applied to the normals shows that a maximum variance plane lies parallel to Saturn's equatorial plane within ~10°. This suggested transforming the normals to a Saturn equatorial system, similar to the δ, λ (heliographic) system, except that Saturn's equatorial plane replaces the Sun's equatorial plane. This requires a 26° rotation about the Saturn–Sun line. Table 1 also shows the transformed method (2′) normals (δ_{eq}, λ_{eq}). The oscillation in the latitude (δ) of the normals has now disappeared, as expected; by comparing method (2) and (2′) normals in Table 1 we note how r.m.s. (δ) has decreased and r.m.s. (λ) has increased. Apparently the surface wave propagation vector is directed along the MP in Saturn's equatorial plane within ≈10°. The transformed estimated normal longitudes (λ_{eq}) are now used to analyse this surface wave[7].

Magnetopause surface wave analysis

The surface wave model assumes that the wave is two-dimensional, transverse, and composed of two partial sinusoids which must match in location, slope, and wave speed at their common point; the wave vector is locally parallel to the zeroth-order MP given by an earlier model[1]. A second assumption is that the observed MP crossings are due entirely to the passing wave, and not to large scale radial motion of the MP. The eight quantities to be estimated are: A_i, the amplitude of the i-part of the wave; α_1, the phase of the i-part; λ_i, the wavelength of the i-part of the wave; V_W, the phase speed of the wave, assumed constant for the full wave; and y_0, the zero-displacement line of the wave, where $i = 1, 2$. The slopes and wave displacements at the two end points of the wave and at the common point between them must match observations, providing eight equations to solve. These three points occur at three successive MP crossings. The input quantities are: V_N, V_T, the components of the velocity of the spacecraft normal and tangential to the zeroth-order MP, respectively, in Saturn's equatorial plane; $\tan\theta$ (where $\theta = \lambda_{eq} - \lambda_0$), the slope of the MP as estimated by the minimum variance analysis; and the times of the three successive MP crossings. The input (centre) times are given in Table 1, $V_N = 16.6$ km s^{-1}, and $V_T = 2.4$ km s^{-1}. The components of the spacecraft's velocity (V_N, V_T) as well as the derived wave speed (V_W) are measured in a system fixed with respect to the Saturn–Sun line. The angle λ_0 is the longitude of the normal to an ideal zeroth-order MP in the locality of the three successive crossings. Three trial λ_0s were used (197°, 200° and 202°) giving three trial sets of $\tan\theta$ for the given λ_{eq}s. The λ_0s are consistent with an estimated λ_0 of 200° ± 3° based on a weighted average of a model[1] normal (λ_0 (model) = 208°) and the average of the derived λ_{eq}s from Table 1 (195°), where the latter was favoured by 2-to-1 in the weighting. Three separate (but not independent) analyses were made of the wave characteristics using: MPs 1, 2 and 3 (Set I); MPs 2, 3 and 4 (Set II); and MPs 3, 4 and 5 (Set III).

Table 2 gives the results of the surface wave calculations, where A_e is the effective wave amplitude based on a weighted average of A_1 and A_2, similarly for the effective wavelength λ_e and $\tau_e = \lambda_e/V_W$; (for the method of weighting see ref. 7). The five trial results not shown ($I_{1,2,3}$ and $II_{2,3}$) were not physically allowable. Regardless of the variability of V_W, A_e and λ_e, the effective period τ_e is consistently 23 (±2) min and λ_e/A_e is nearly constant at 10 (±1). The latter fact indicates that the wave displacement is apparently of relatively small amplitude.

We now check the assumption that the MP crossings were primarily due to the surface wave and not to bulk normal motion of the MP. The time between centres of the first and last crossings is 52 min. Then $\langle A_e\rangle \approx \Delta T \times V_N/2$, which is 26,000 km; the wave analysis gives a very favourable comparison of $\langle A_e\rangle = 25{,}600$ km (Table 2).

Simple geometrical considerations give an MP thickness

$$D = |\Delta t[(V_W - V_T)\sin(\lambda_{eq} - \lambda_0) + V_N \cos(\lambda_{eq} - \lambda_0)]| \quad (1)$$

for each crossing. All quantities in equation (1) are fairly accurately estimated, measured, or known, except λ_0 and V_W, which were moderately well estimated. We assume V_W to be 180 km s^{-1}, the average of the four estimated cases (Table 2). The Δts and λ_{eq}s are given in Table 1. Consistent with our choice

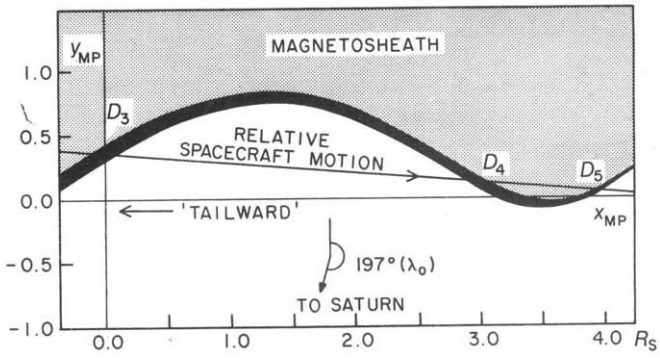


Fig. 2 An average example (case III_1) of Saturn's magnetopause surface waves. The unperturbed magnetopause is aligned with the X_{MP}-axis. $D_{3,4,5}$ refer to magnetopause thicknesses.

© 1981 Nature Publishing Group

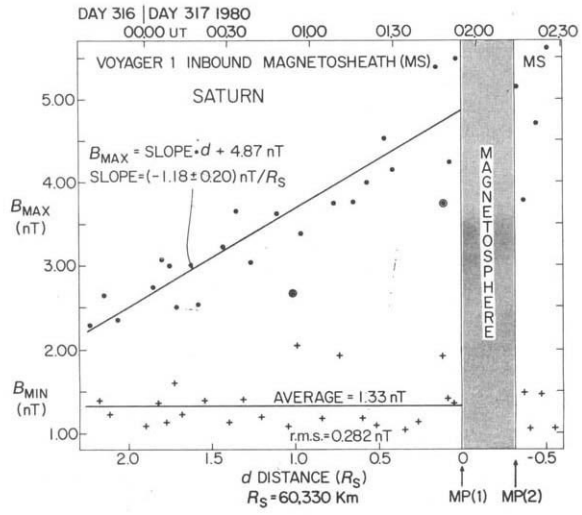


Fig. 3 The relative maximum (B_{MAX}, •) and minimum (B_{MIN}, +) magnetic field magnitudes, based on 9.6-s averages, for the inbound magnetosheath regions, as a function of distance (d) from the first magnetopause crossing and time. ●, points not used in the straight line fit of B_{MAX}.

of V_W we assume that $\lambda_0 = 197°$ and 200°, those giving intermediate wave characteristics. The resulting estimated MP thicknesses range from 2,300 km (crossing 5) to 8,500 km (crossing 2), giving $\langle D \rangle = 5.0 \pm 3.0 \times 10^3$ km, averaged over all five crossings and over both λ_0s; $\lambda_0 = 197°$ gave the smaller spread of values. Note that $\langle A_e \rangle$ is about five times greater than $\langle D \rangle$, for either λ_0, consistent with another implicit assumption that the wave amplitude is significantly greater than the MP thickness.

Figure 2 shows the MP surface wave derived from trial III_1 (that is $\lambda_0 = 197°$) chosen because it most closely resembles the average of all successful trials (Table 2). The variable MP thickness shown is based on $\lambda_0 = 197°$ for crossings 3, 4 and 5. Such surface waves at this location on the MP ($\approx$ 13.10 LT) near Saturn's equatorial plane are possibly due to a Kelvin–Helmholtz instability[9] at the boundary, on either side of which plasmas are flowing parallel to it but at significantly different speeds. The difference in speed across the MP in Saturn's equatorial plane must be closely comparable with the wave speed $\langle V_W \rangle$, that is

$$\langle V_W \rangle \approx V(\text{sphere}) - V(\text{sheath}) \qquad (2)$$

if the Kelvin–Helmholtz instability is the causal mechanism. In our case the magnetosphere plasma speed, V(sphere), is faster than the magnetosheath plasma speed, V(sheath). Using an Earth-like magnetosheath model[10] for $M_\infty = 12$ and $\gamma = 5/3$ at 13.10 LT and a typical solar wind speed of 400 km s^{-1} (close to that actually measured for many hours before the inbound bow shock (J. Scudder, personal communication)), gives V(sheath) = $0.16 \times 400 = 64$ km s^{-1}; this estimated magnetosheath speed is not very sensitive to M_∞, provided it is $\geqslant 5$. M_∞ was calculated to be ~13–15 where the upstream proton and electron temperatures ($T_p \approx 2.3 \times 10^4$ K and $T_e \approx 1.5 \times 10^4$ K) were provided by the Voyager Plasma Science Team. For $\langle V_W \rangle = 180$ km s^{-1} and this magnetosheath speed, equation (2) yields V(sphere) $\approx$ 240 km s^{-1}. If Saturn's magnetosphere rigidly corotates out to the frontside MP at $\approx 23\ R_S$ (where $R_S \equiv 60{,}330$ km, Saturn's radius) at this time, the corotation speed at the MP would be 230 km s^{-1}, close to the above result. As the uncertainty on our estimate of V(sphere) is probably quite large, we cannot say that rigid corotation of the magnetosphere holds to 23 R_S, but only that this important necessary condition (equation (2)) for showing that the Kelvin–Helmholtz instability may be applicable is not obviously violated. The fact that the MP surface normals oscillate in Saturn's equatorial plane, rather than in a plane containing the local magnetosheath plasma flow ($\approx$ the ecliptic plane), supports the view that it is primarily the magnetosphere motion which drives the instability.

Table 2 Surface wave characteristics

Quantity	($\lambda_0 = 197°$) II_1	(197°) III_1	(200°) III_2	(202°) III_3	Average
V_W (km s^{-1})	346	210	90	68	180
A_e ($\times 10^4$ km)	4.97	2.70	1.37	1.18	2.56
A_e [$\times R_s$]	[0.824]	[0.448]	[0.227]	[0.196]	[0.424]
λ_e ($\times 10^5$ km)	5.06	2.80	1.30	1.02	2.55
λ_e [$\times R_s$]	[8.39]	[4.64]	[2.15]	[1.69]	[4.22]
τ_e (min)	24	21	24	25	23.5
λ_e/A_e	10.2	10.4	9.5	8.6	9.7

Inbound magnetosheath waves

Here we briefly discuss the oscillations occurring in the inbound magnetosheath field, mainly in its magnitude, B, as shown in Fig. 1; these have an average period of ~6 min. as measured in the spacecraft frame of reference. The latitude ($\langle \delta_B \rangle \approx 0°$) and longitude ($\langle \lambda_B \rangle \approx 285°$) of the magnetosheath field agrees with an expected wrapped-up solar wind field at Saturn's orbit ($\lambda_B \approx$ 270° or 90°). Oscillations in field direction, especially in longitude, are present, but dominated by the magnitude changes. The values of the local maxima and minima of the field in the magnetosheath are plotted in Fig. 3 against normal distance from the first MP(1), assuming a stationary zeroth-order MP at this time and using the spacecraft speed normal to the MP of $V_N = 16.6$ km s^{-1}. Straight line fits were made to the B_{MAX} and B_{MIN} points; the B_{MIN} set was assumed to have zero slope. The wave has apparent amplitude (A) growth given by

$$A = (B_{MAX} - B_{MIN})/2 = \text{slope} \times d/2 - 1.77 \text{ nT} \qquad (3)$$

where slope = (-1.18 ± 0.20) nT/R_S according to the straight-line fits. The marked growth in amplitude as the MP is approached suggests that the MP may be the source of these waves; notice the highest B_{MAX} amplitudes were maintained in the magnetosheath even when the spacecraft returned there between $\approx$02.15 and 02.30 UT. Similar waves in the inbound Pioneer 11 magnetosheath data were reported by Smith *et al.*[3]. The Pioneer observations, however, differed in one respect: the B_{MIN}s did not remain nearly constant but grew approximately in line with the B_{MAX}s, yielding $A \approx$ constant. The apparent difference is either a different phenomenon or simply due to different conditions, such as possibly an outward moving MP in the Pioneer 11 case and a static (excluding surface waves) MP for Voyager. In any case, a gradual increase in B is probably not needed to produce the waves.

Examination of the Voyager plasma science electron data (E. Sittler, personal communication) for the period shown in Fig. 3 indicates that the electron density was ~180° out-of-phase with B for these waves, suggesting that they are slow mode magnetosonic waves, which have their most rapid phase velocity along the ambient magnetic field. If the MP is the source of these waves, they must have a component of their propagation vector upstream, and it must be greater than the incoming convected magnetosheath plasma velocity; this is possible, because at this location the magnetosheath plasma is subsonic. Although the waves in B resemble solitons[3], β in the magnetosheath is large, so a large perturbation in B does not necessarily imply a correspondingly large change in density.

Summary and discussion

Voyager 1 magnetometer data have revealed that surface waves occur on Saturn's dayside MP. Estimates of Saturn's magnetopause wave properties show that: the wave speed is 180 ± 90 km s^{-1}, amplitude $12\text{–}50 \times 10^3$ km (or ~0.5 R_S), wavelength $1\text{–}5 \times 10^5$ km (or ~5 R_S), and a well determined period 23 ± 2 min. Most estimates yielded a consistent ratio of wavelength-to-amplitude of 10 ± 1, and a typical magnetopause thickness of $5.0 \pm 3.0 \times 10^3$ km or ~0.1 R_S was determined.

The success of the model was partly due to the large-scale stationarity of the MP during the analysis period, probably the result of apparently persisting quiet upstream solar wind conditions, which was the case before the inbound bow shock cross-

© 1981 Nature Publishing Group

ing. The occurrence of only a single bow shock crossing shortly before the multiple MP crossings is consistent with our basic premise that the surface waves caused the multiple crossings at a time when the solar wind was fortuitously quiet.

Hydromagnetic waves occurring in the inbound magnetosheath and having an average period in the spacecraft frame of reference of 6 min grew markedly in amplitude as the magnetopause was approached. This suggests the MP as a possible source of these waves unless the amplitude growth is due to a geometrical effect. These waves have been tentatively identified as slow mode magnetosonic waves. Recently such waves have been observed in the Earth's magnetosheath by the ISEE spacecraft (J. Scudder, personal communication).

We thank our Voyager colleagues for helpful comments, and especially thank J. E. P. Connerney, M. L. Kaiser, W. H. Mish, N. F. Ness, J. D. Scudder, E. C. Sittler and B. T. Tsurutani.

Received 13 May; accepted 13 July 1981.

1. Ness, N. F. *et al. Science* **212**, 211–217 (1981).
2. Behannon, K. W. *et al. Space Sci. Rev.* **21**, 235–257 (1977).
3. Smith, E. J. *et al. Science* **207**, 407–410 (1980); *J. geophys. Res.* **85**, 5655–5674 (1980).
4. Sonnerup, B. U. Ö. & Cahill, L. J. *J. geophys. Res.* **72**, 171–183 (1967).
5. Siscoe, G. L., Davis, L. Jr, Coleman, P. J. Jr, Smith, E. J. & Jones, D. E. *J. geophys. Res.* **73**, 61–82 (1968).
6. Burlaga, L. F., Lemaire, J. & Turner, J. M. *J. geophys. Res.* **82**, 3191–3200 (1977).
7. Lepping, R. P. & Burlaga, L. F. *J. geophys. Res.* **84**, 7099–7106 (1979).
8. Fairfield, D. H. in *Magnetospheric Particles and Fields* (ed. McCormac, B. M.) 67–77 (Reidel, Hingham, 1976).
9. Lee, L. C., Albano, R. K. & Kan, J. R. *J. geophys. Res.* **86**, 54–58 (1981).
10. Spreiter, J. R., Alksne, A. Y. & Summers, A. L. *NASA TN D-4482* (National Aeronautics and Space Administration, Washington DC, 1968).

Saturn's magnetic tail: structure and dynamics

K. W. Behannon, J. E. P. Connerney & N. F. Ness

Laboratory for Extraterrestrial Physics, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Voyager 1 magnetic field observations have provided evidence of a saturnian magnetic tail. Tail current system distributions are inferred through comparison of the observations with a realistic magnetotail current system model. Temporal variations observed in the tail were probably produced by solar wind variations.

SATURN'S magnetic tail was first observed by Voyager 1 as it tranversed the saturnian magnetosphere outbound at 03.40 LT. The observed magnetic field directions clearly indicated the sweeping back of field lines into a tail configuration pointing away from the Sun[1] and led us to conclude that Saturn, like Earth and Jupiter, possesses a magnetotail that develops in the interaction of the solar wind with the planetary magnetic field. Closest approach for Voyager 1 at Saturn occurred at 23.45 UT on 12 November 1981. Between 17.29 and 21.52 UT on 14 November, it crossed the magnetopause outbound five times at an average distance from Saturn of 45 R_S (24 R_S antisolar distance, where R_S = Saturn radii = 60,330 km) and at a latitude of ~24° (ref. 1). The encounter trajectory characteristics, together with the length of time (~45 h) spent in the night-side magnetosphere, gave ample opportunity to study the geometry of the tail. We estimated the tail diameter to be ~80 R_S, less than that of the jovian tail by a factor of 4–6. From the observed size of the tail and the average magnitude of the distant tail lobe field that was measured (~3 nT), a conservation of flux argument gave an estimate of 11°–15° for the half-angular size of the saturnian polar cap auroral zone, smaller than that of Earth[1]. Because of the high latitude of the outbound pass, there were no traversals of the tail current sheet. Only the northern lobe field (directed away from the planet) was measured. The near alignment of Saturn's magnetic dipole and rotational axes made the outbound traversal a nearly constant magnetic latitude path through a planetary magnetosphere.

Although there was no direct tail current sheet observation, plausible current distributions and geometry can be inferred through comparison of measurements with the fields calculated from simple but realistic magnetosphere models that include magnetotail current systems. The preliminary results of the study of one such model are described below.

Model magnetosphere

We have studied a model of Saturn's magnetosphere which is consistent with the Voyager 1 magnetic field observations but which must also be non-unique. A noon–midnight meridian projection of the model field and the Voyager 1 outbound hourly averaged magnetic field observations is shown in Fig. 1. The closure of the tail field lines is very sensitive to the detailed current distribution in the more distant (and unexplored) tail and is not a direct result of the fit to the magnetic field observations. It is, rather, a boundary condition imposed on this model determined, in part, by the observed electron pitch angle distributions[2] which are indicative of closed field lines. The closure of field lines in the tail is uncertain. Observations of solar cosmic rays to 9 R_S by the Pioneer 11 cosmic ray experiment[3] suggest open field lines in the magnetotail, although Saturn's magnetosphere was in a disturbed state during the Pioneer 11 encounter.

The model field has the following components: an internal dipole field of 0.21 G$-R_S^3$ aligned with z_{SM}; a distributed near-equatorial disk or ring current in the region $8<\rho<16$ R_S; cross-tail currents extending from $x_{SM}=-16$ to -100 R_S, closing on the magnetopause boundary; and confinement of the day-side magnetosphere by image dipoles. The azimuthal ring current at Saturn is modelled on a similar structure at Jupiter[4] and will not be discussed here. The field of the cross tail currents

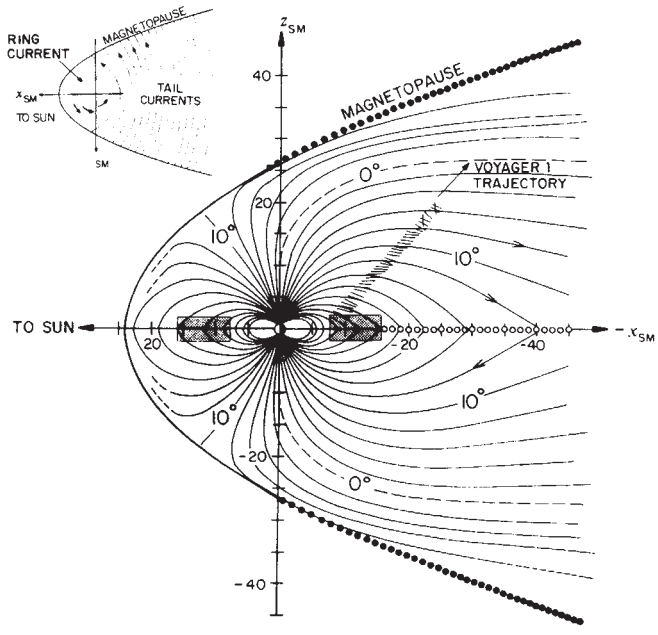


Fig. 1 Conceptual and quantitative model of Saturn's magnetosphere in the noon–midnight meridian plane. Field lines are drawn for 2° increments of invariant latitude. Azimuthal disk or ring current (stippled) is indicated, as well as model cross-tail currents closing on the magnetopause boundary. Projections of Voyager 1 outbound hourly-averaged magnetic field observations onto the solar magnetospheric (SM) x–z plane are also shown. Insert, model cross-tail currents in the SM x–y plane. Current flows eastward in parabolic paths to the magnetopause (MP) pause boundary and closes in planes of constant x equally on both the northern and southern lobe MP surfaces.

0028–0836/81/340753—03$01.00

© 1981 Macmillan Journals Ltd

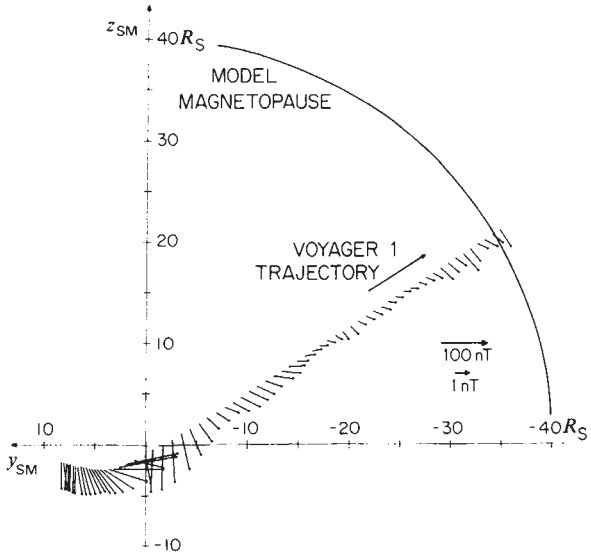


Fig. 2 Projection of hourly-averaged magnetic field components in the saturnian magnetosphere onto the SM y–z plane along the projected Voyager 1 trajectory. A logarithmic scale has been used to represent field magnitude as indicated.

is computed numerically by summation of the field due to a system of individual closed current loops similar to that used in models of the Earth's magnetosphere[5,6]. Individual loops are spaced at $1-R_S$ intervals and carry an eastward current in the solar magnetospheric (SM) equatorial plane as indicated in Fig. 1. The SM coordinate system is a right-handed, orthogonal, non-rotating system defined such that $\hat{x}_{SM}$ is directed from the planet to the Sun and the $\hat{z}_{SM}$ axis lies in the plane formed by $\hat{x}_{SM}$ and $\vec{M}$, the magnetic dipole of the planetary field. The current has a $1/(x_{SM})^{1/2}$ dependence, decreasing from ~5 A km^{-1} at x_{SM} of $-16\ R_S$, which is about half the (linear) current density of the distributed ring currents at their outer edge.

In this view of Saturn's magnetic tail, the high latitude field lines observed along most of the Voyager 1 trajectory remain closed, while the observed magnetic field orientation along the trajectory remains essentially parallel to the SM equatorial plane. The estimated size of the polar cap[1] based on the conservation of flux using an estimate of 3 nT for the tail field at $x_{SM} = -25$ may represent an upper limit on the size of the polar cap. As indicated in Fig. 1, to the extent that field lines cross the equatorial plane at greater $-x_{SM}$, Saturn's polar cap may be smaller and is not symmetric with respect to the dusk meridian.

A consequence of the confinement of Saturn's magnetosphere to a parabolic magnetopause boundary[1] is the very Earth-like day-side cusp at ~80° invariant latitude. The inferred source region of Saturn kilometric radiation (SKR) is centred about the intersection of the noon–midnight meridian and 80° invariant latitude[7]. This suggests that SKR is related to particle precipitation in the cusp, analagous to the terrestrial kilometric radiation associated with the Earth's polar cusp[8].

Tail dynamics

Variations of the magnetotail magnetic field were observed by Voyager 1 during the outbound traversal[1]; they appear as moderately small amplitude, long-period excursions in field direction and magnitude. The directional variations can be seen in the hourly average SM y–z plane vector projections of Fig. 2. The x–y and x–z plane projections have been presented previously[1]. Assuming that the observed changes are most likely temporal and not spatial, is the source of these perturbations of the tail lobe field internal or external to the magnetosphere? Quasiperiodic variations of the flux of low energy charged particles were also observed during this period[2], and a Saturn-related magnetic anomaly-type feature suggested as the cause.

The magnetic field measurements have been analysed to try to elucidate the source location of the perturbations producing the observed variations. To test the possibility of an internal, planet-associated (and hence periodic) source, hourly values of the magnetic field direction angles λ and δ have been plotted against sub-spacecraft Saturn longitude in superposed epochs for the 36-h period when Voyager 1 was at a radial distance $r > 10\ R_S$ (Fig. 3). Here λ and δ are the SM azimuth and latitude angles, respectively, of the field. Because the dipole tilt relative to the rotation axis is ~1°, SM coordinates are defined for Saturn such that λ is measured anticlockwise in the magnetic dipole equatorial plane, as viewed from north of (above) the plane, with the $\lambda = 0°$ meridian plane at local noon.

The data suggest a recurrence tendency at roughly the planetary rotation rate, but with changes evident in the amplitude and phase of the variation. Note that Voyager 1 was in the tail lobe for only ~2 − 1/2 planetary rotation periods, and it is therefore difficult to establish a recurrence pattern. If recurrence is assumed, a period of 12.1 h is obtained for the field variation, significantly greater than the 10.66 h planetary rotation period deduced by the Voyager 1 Planetary Radio Astronomy experiment[9]. The longer period seen at Voyager 1 may be caused by Doppler-shifting of the perturbation signal as it 'propagates' like a wave from Saturn. Taking into account the velocity of Voyager 1 relative to Saturn, a radial wave speed $V_r \simeq 140$ km^{-1} is obtained.

This wave speed can be tested for consistency with the Alfvén speed $V_A = B/(4\pi mn)^{1/2}$ expected in the tail lobe plasma. The ion number density, n, is assumed to be equal to that reported for the plasma electrons[10]. At distances $r = 10\ R_S$ and $40\ R_S$, respectively, field magnitudes of 21 and 3 nT and densities of 0.3 and 0.003 cm^{-3} were measured. Assuming the ions are protons gives $V_A = 836$ km s^{-1} in the near tail, increasing to 1,196 km s^{-1} in the distant tail. If the ions were nitrogen, these values would decrease to 316 and 452 km s^{-1}, respectively. The V_A values for a nitrogen plasma are factors of more than 2–3 higher than the Doppler-shift predicted wave speed of 140 km s^{-1} and factors of 6–8 too high in the case of hydrogen. Conversely, assuming $V_A = 140$ km s^{-1} predicts densities for hydrogen and nitrogen of $n_H = 10.7$ cm^{-3} and $n_N = 1.5$ cm^{-3} at $r = 10\ R_S$ compared with the observed $n_e = 0.3$ cm^{-3}, and $n_H = 0.22$ cm^{-3} and $n_N = 0.03$ cm^{-3} at $r = 40\ R_S$ compared with $n_e = 0.003$ cm^{-3}.

These results indicate an inconsistency between the observations and the hypothesis of a planet-associated propagating wave. Furthermore, because the magnetic dipole axis tilt relative to the axis of rotation is estimated to be in the order of only 1°, the dipole wobble is unlikely to produce the observed variations.

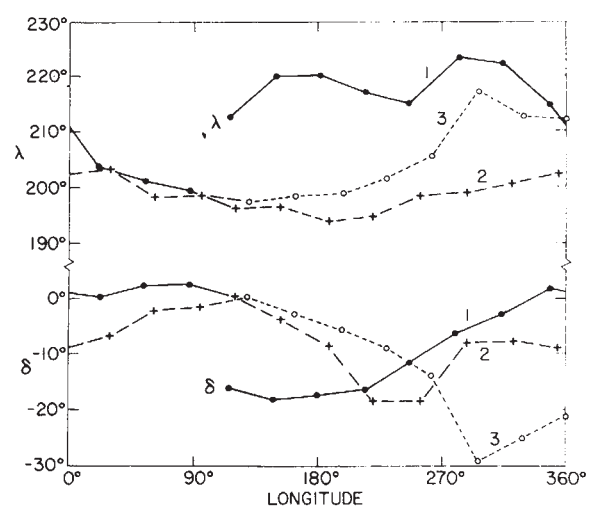


Fig. 3 Superposed epoch presentation of hourly values of magnetotail magnetic field azimuthal (λ) and latitude (δ) angles as functions of planetary longitude for three successive rotations of the planet (1, 2 and 3). The data do not demonstrate clearly a longitude-based recurrence pattern.

© 1981 Nature Publishing Group

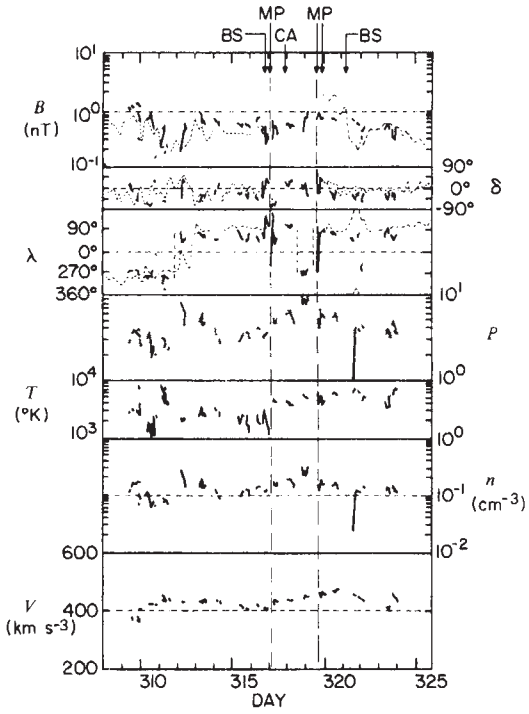


Fig. 4 Predicted solar wind parameters and interplanetary magnetic field during the Voyager 1 encounter period. These data were derived from the solar wind measurements by Voyager 2 upstream of Saturn (see text). Plasma parameters shown include bulk speed (V); number density (n); temperature (T); and plasma pressure (P). The magnetic field is given in terms of its magnitude, B, and the heliographic latitude and azimuthal angles, δ and λ. Actual Voyager 1 measurements are also given (dashed curves). Voyager 1 magnetopause (MP) crossing times are given by arrows at top and further delineated by vertical lines. Bow shock (BS) crossings and closest approach (CA) times are also indicated by arrows at the top.

An investigation of the solar wind during the encounter period indicates that perturbations outside the saturnian magnetosphere may have been the source of the observed variations. The solar wind and interplanetary magnetic field (IMF) conditions at Saturn during the encounter have been predicted by applying a transformation to Voyager 2 interplanetary hourly average field and plasma observations (J. D. Sullivan and H. S. Bridge, personal communication). This projects the observations in two-dimensional space and in time assuming that each parcel of plasma expands radially outward at constant speed. The hourly averaged data propagated in this way from Voyager 2 to Saturn are shown in Fig. 4. Time is given as day-of-year number (day 317 = 12 November). The distorted and fragmented appearance of the data is due to effects of the variable speed correction applied in time-shifting each parcel and the intermittent tracking and data retrieval from Voyager 2. The sizeable gaps (~14 h) mean that a one-to-one correlation between changes in these data and those seen by Voyager 1 inside the saturnian magnetosphere is impossible. However, the data show substantial changes in the solar wind and IMF during the encounter. The lighter dashed data curves superimposed on the predicted data in the top three panels are the actual Voyager 1 IMF observations for comparison. The field data are in spacecraft-centred heliographic coordinates, a spherical system with azimuth $\lambda = \tan^{-1}(B_T/B_R)$ and $\delta = \sin^{-1}(B_N/B)$, where $\hat{R}$ is radially away from the Sun; $\hat{T}$ is perpendicular to $\hat{R}$ and parallel to the solar equatorial plane, positive in the general direction of Saturn's orbital motion; and $\hat{N} = \hat{R} \times \hat{T}$.

An examination of λ shows a prediction of an IMF sector transition occurring ~6 days before closest approach (CA) and a similar, return transition occurring between 7 and 16 h after CA. These are accentuated by vertical dashed lines at the midpoints of the gaps in the λ data where the transitions occurred. In the latter transition, which would have taken place while Voyager 1 was in the tail, the new state of the medium only persisted for between 10 and 32 h and thus was either a very narrow sector or a filamentary-type structure. The field magnitude increased within this region, accompanied by a plasma density increase of about a factor of two and a corresponding increase in solar wind pressure. The leading edge of this region is predicted to have passed between the middle and the end of day 318—the period when the temporal excursions in the magnetotail began.

The major uncertainty is that the Voyager 2 data are being extrapolated over a distance of ~1.7 AU, and in that distance the characteristics of plasma streams and associated IMF structures can be modified significantly by dynamical processes. A comparison of the projected Voyager 2 data with measurements from Voyager 1 before and after passing through Saturn's magnetosphere (included in Fig. 4) confirms the essential features of the prediction. However, the sector transition predicted to occur at the end of day 311 was actually observed to begin ~12 h later. A similar shift in the onset of the perturbations during tail passage would start their effect at a later time but with Voyager 1 still in the magnetotail. Different transformation conditions (higher solar wind speed) could have produced a closer agreement in this case between predicted and observed features than was found during the earlier, pre-encounter period.

While the direct consequences for a magnetotail orientation and field strength of changes in external solar wind direction and pressure on the boundary are generally understood, how changes in the IMF configuration influence a magnetotail is not completely understood. Studies of the terrestrial magnetosphere show that this influence is probably exerted in most cases indirectly through a chain of events controlled by the IMF-to-magnetospheric field reconnection efficiency and including the process of flux transfer within the magnetosphere[11,12]. Because the observations at Saturn are limited to one tail lobe, and because of the supporting interplanetary measurements, we can show only that plausible conditions existed in the interplanetary medium for stimulating the observed tail perturbations, with the precise mechanism remaining a matter of speculation. As an internal source mechanism for the perturbations seems inconsistent with measured physical parameters, we conclude that an external source, the solar wind, is responsible.

We thank our Voyager plasma investigation colleagues at MIT for providing the projected Voyager 2 plasma and magnetic field data for use as a prediction of solar wind conditions at Saturn during the Voyager 1 encounter.

Received 13 May; accepted 10 June 1981.

1. Ness, N. F. *et al. Science* **212**, 211–217 (1981).
2. Krimigis, S. M. *et al. Science* **212**, 225–231 (1981).
3. McDonald, F. B., Schardt, A. W. & Trainor, J. H. *J. geophys. Res.* **85**, 5813–5830 (1980).
4. Connerney, J. E. P., Acuña, M. H. & Ness, N. F. *J. geophys. Res.* **86** (in the press); *Nature* **292**, 724–726 (1981).
5. Olson, W. P. *J. geophys. Res.* **79**, 3731–3738 (1974).
6. Walker, R. J. in *Quantitative Modeling of Magnetospheric Processes* ed. Olson, W. P. 9–34 (American Geophysical Union, Washington DC, 1979).
7. Kaiser, M. L., Desch, M. D. & Lecacheux, A. *Nature* **292**, 731–733 (1981).
8. Alexander, J. K., & Kaiser, M. L. *J. geophys. Res.* **82**, 98–104 (1977).
9. Desch, M. D. & Kaiser, M. L. *Geophys. Res. Lett.* **8**, 251–256 (1981).
10. Bridge, H. S. *et al. Geophys. Res. Lett.* **8**, 217–224 (1981).
11. Holzer, R. E. & Slavin, J. A. *J. geophys. Res.* **84**, 2573–2578 (1979).
12. Holzer, R. E. & Slavin, J. A. *J. geophys. Res.* **86**, 675–680 (1981).

© 1981 Nature Publishing Group

ARTICLE

Total synthesis of a human leukocyte interferon gene

Michael D. Edge, April R. Greene, Gillian R. Heathcliffe, Peter A. Meacock*, Wolfgang Schuch†, Denis B. Scanlon, Thomas C. Atkinson, Clive R. Newton & Alexander F. Markham

ICI Pharmaceuticals Division, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK
* ICI/University Joint Laboratory, School of Biological Sciences, University of Leicester, Leicester LE1 7RH, UK
† ICI Corporate Laboratory, The Heath, Runcorn, Cheshire WA7 4QE, UK

A 514-base pair fragment of double-stranded DNA coding for human interferon-α_1 (166 amino acid residues), and containing initiation and termination signals plus appropriate restriction enzyme sites for plasmid insertion, has been totally synthesized. The synthesis involved preparation of 66 oligodeoxyribonucleotides, ranging in size from 14 to 21 residues, plus 1 deoxydecanucleotide, by rapid, solid phase procedures, and enzymatic ligation of the oligonucleotides. After ligation of the synthetic gene to a plasmid vector and transformation of Escherichia coli, *clones containing the anticipated gene sequence were obtained.*

THE antiviral[1,2], cell proliferation inhibition[3] and immunomodulating properties[4] of human interferons (IFNs) have generated great interest because of their potential clinical value in the treatment of viral infections and malignancies. Thorough investigation of the efficacy of both leukocytic (IFN-α) or fibroblastic (IFN-β) interferons has, however, been hampered by an inability to purify large amounts of these proteins from human cells.

The recent application of recombinant DNA techniques has dramatically changed this situation. Several groups have obtained cDNA clones prepared using mRNA from a variety of human cells induced to synthesize interferon(s)[5-10]. DNA sequencing has allowed the amino acid sequences of a number of IFN-αs[11] and IFN-β[12,13] to be deduced (as well as their putative signal sequences) and various elegant genetic manipulations have generated expression plasmids which direct the bacterial synthesis of biologically active interferons[14-21]. A second phase of this research has seen eight or more distinct human leukocyte interferon genes[22] identified in human DNA libraries and neither these IFN-α genes[23,24], nor the IFN-β gene[25,26], apparently contain introns. In some of the above studies the availability of specific oligonucleotides, designed on the basis of limited amino acid sequence data[27,28], was invaluable for the identification of cloned IFN cDNA. Furthermore, synthetic oligonucleotides have been generally useful in the various gene editing procedures used for the construction of expression plasmids.

On the basis of the nucleotide sequence of the first cloned human IFN-α cDNA published by Weissmann and co-workers[7] (IFN-α_1), we designed a DNA fragment which would encode the same protein and be amenable to total chemical synthesis (Fig. 1). Such a synthetic gene has several potential advantages, which are discussed further below. However, it is obviously desirable that such a synthesis be accomplished in a comparable time scale with that required for isolation of the gene from natural sources. Various other synthetic genes, including those for human somatostatin[29], human insulin A and B chains[30] and N^{α}-desacetylthymosin α_1 (ref. 31), have previously been constructed but these molecules are all considerably smaller than the IFNs. Total synthesis of an IFN gene would have been unrealistic using standard solution phase phosphotriester methodology as applied in the above examples[32]. Recently, we have developed a novel solid phase phosphotriester approach for the synthesis of oligonucleotides[33,34]. This increases the rate of synthesis by an order of magnitude and is routinely reliable, the inherent hazards of repeated chromatographic purification being largely avoided.

Using these methods, 67 different oligonucleotides containing around 15 nucleotide residues were synthesized. They were separately purified and 5′ phosphorylated, then sequentially ligated to yield a 514-base pair (bp) fragment. After insertion in a hybrid plasmid and cloning in *Escherichia coli*, the synthetic

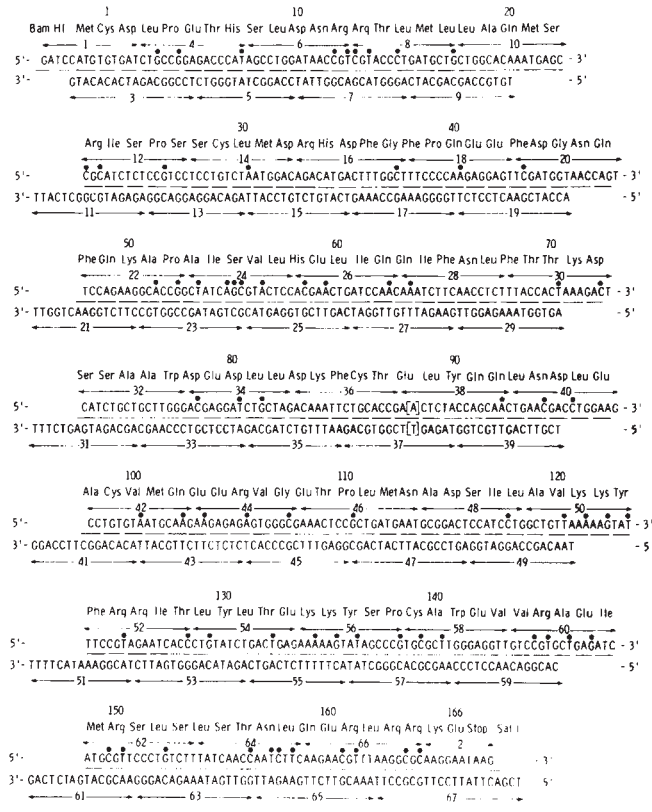


Fig. 1 Nucleotide and amino acid sequence of the synthetic human leukocyte interferon gene. Individual oligonucleotides (1–67) are arrowed. Dots above bases in the coding strand indicate changes with respect to the native sequence[7]. Numbers above each line refer to amino acid position.

0028–0836/81/340756—07$01.00

© 1981 Macmillan Journals Ltd

gene was characterized by restriction enzyme digests and total sequence determination. Biological activity of material expressed by these clones will be described elsewhere (E. De Maeyer, manuscript in preparation).

Design of the synthetic gene

Initially, the published DNA sequence[7] coding for mature human IFN-α_1 was scanned with a computer program designed to detect both complementary and repeated sequences (D. J. Gilman, unpublished results). A number of such sequences were identified. For example, bases 75–85 in the native sequence, d(TCCTTCCTCCT), are complementary to bases 487–497, d(AGGAGGAAGGA), and bases 32–40, d(ACAGGAGGA) are complementary to bases 79–87, d(TCCTCCTGT). Similarly, there are 4 separate complementary pairs of octamers and 13 complementary pairs of heptamers. The sequence d(GAAGAAATAC) (bases 359–368) is repeated (bases 399–408). There are seven separate repeated heptamers, five repeated octamers and one repeated nonamer. Numbering here is with respect to the first base in the triplet coding for N-terminal cysteine (T = 1) in the mature peptide (see Fig. 3 in ref. 7).

We reasoned that elimination of these features as far as possible within the constraints of the genetic code would make correct ligation of synthetic oligonucleotides more straightforward. The target sequence is illustrated in Fig. 1. Dots above bases in the coding strand indicate changes with respect to the native sequence (71 in total). By so modifying the natural sequence, the longest self-complementary regions are four heptamers, all separated by more than 100 bp. There are two repeated nonanucleotides; d(AATCTTCAA) in blocks 28 and 64/66 and d(AAAAAGTAT) in blocks 50 and 56. There are three repeated heptamers and one repeated octamer.

The gene was divided (Fig. 1) into a series of oligonucleotides of ~15 residues so as to give overlaps of at least 7 bp on each side of all ligation points. It is convenient to use oligonucleotides of this size as the synthesis and purification of significantly larger molecules may not always be straightforward[33]. Where possible, changes in the native sequence are such as to introduce codons which may be preferred for high levels of expression in bacteria[35–37]. The codon usage profile is compared with that of the native strand in Table 1.

Outline of chemical synthesis strategies

The basic chemistry involved in our synthetic approach is summarized at the top of Fig. 2 and in its legend and we have previously described the methodology in some detail[33,34]. During this study several minor changes have been made which considerably improve both the rate and yields of synthesis. In our hands, use of mesitylenesulphonyl-3-nitro-1,2,4-triazole[38] as condensing agent increases yields in dimer coupling reactions and allows individual reaction times to be reduced to ~1 h. We have also reduced the deprotection/washing cycle to ~0.5 h so that the overall time for addition of each dimer block to a growing chain is ~1.5 h. With a semi-automated solvent delivery system it is convenient to perform two or three such assemblies concurrently so that up to two pentadecanucleotides can be constructed per man-day. Previously we have not attempted to dry the support between coupling cycles but co-evaporation of the polydimethylacrylamide resin with anhydrous pyridine before each condensation reaction again significantly improves the final yields of purified deprotected oligonucleotides. Thus (Fig. 2*a*, *c*), in the syntheses of blocks 24 and 65 (Fig. 1) the desired oligonucleotide product is the major peak in ion-exchange HPLC on Partisil 10-SAX.

Further resolution by reverse-phase HPLC on μ-Bondapak C_{18} columns (Fig. 2*b*, *d*) gives pure material in most cases, as judged by standard 5′-[^{32}P]-phosphorylation and sizing on 20% polyacrylamide gels in 7 M urea (for examples see Fig. 2*e*, *f*)[39]. In our previous work[33,34], oligonucleotides synthesized by this approach were characterized by complete sequencing. During this synthesis, the oligonucleotides were characterized by sizing (above) and 5′ residues were confirmed after total snake venom phosphodiesterase digestion of 5′-[^{32}P]-labelled aliquots. In cases where unequivocal identification of product by HPLC and sizing was not possible, complete sequencing was undertaken.

A more complete description of improved methodology for the synthesis of these oligonucleotides will be published elsewhere. The oligomers 5, 7, 17, 44, 58 and 63 could not be purified by standard Partisil 10-SAX ion-exchange HPLC with phosphate gradients in a 5% ethanol system at ambient temperature. Elevated temperature and increased ethanol concentration or electrophoretic techniques were necessary[33]. We note that all these oligomers contain three or more consecutive dG residues. In retrospect, these features could have been avoided in oligomers 5, 7, 17, 44 and 63 by alternative choices of codons but it is obviously desirable to be able to synthesize such molecules so as to have the widest possible latitude in codon usage. Furthermore, this sequence cannot be avoided in oligomer 58 given the amino acid sequence Trp-Glu [d(TGGGAPur.)].

Ligation strategy

Eleven groups of oligomers (A–K) were initially ligated separately (Fig. 3*a*). 5′ Phosphorylation with low specific activity ^{32}P-ATP and polynucleotide kinase was performed on individual oligomers which were then isolated by elution from 20% polyacrylamide gels in 7 M urea (oligomers 1 and 67 were not phosphorylated at this stage). After denaturation, slow annealing and ligation, reaction mixtures were examined on polyacrylamide gels in both native and denaturing conditions (Fig. 3*b*, *c*). Ligated products of the expected size were isolated by preparative polyacrylamide gel electrophoresis in non-denaturing conditions, elution and ethanol precipitation. Product recovery was estimated by counting aliquots dried separately on GFC filter disks. Aliquots of the 11 fragments were combined, annealed and again ligated with T4-induced DNA ligase to give the four blocks shown in Fig. 3*d*. Ligation of these four blocks in a similar manner gave a discrete band of approximately the expected size as identified by 3.5% polyacrylamide native electrophoresis with appropriately sized markers (Fig. 3*e*). This fragment was isolated either by excision and elution from the gel or by gel permeation chromatography of the reaction mixture. 5′ Ends were phosphorylated by standard techniques for plasmid insertion.

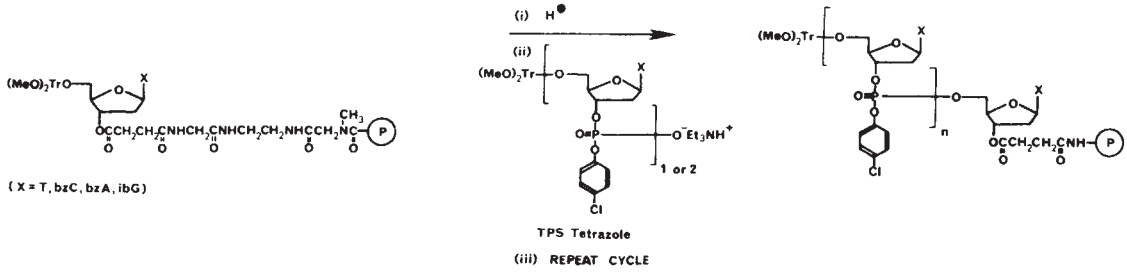
Table 1 Codon usage in synthetic (and natural) human IFN-α

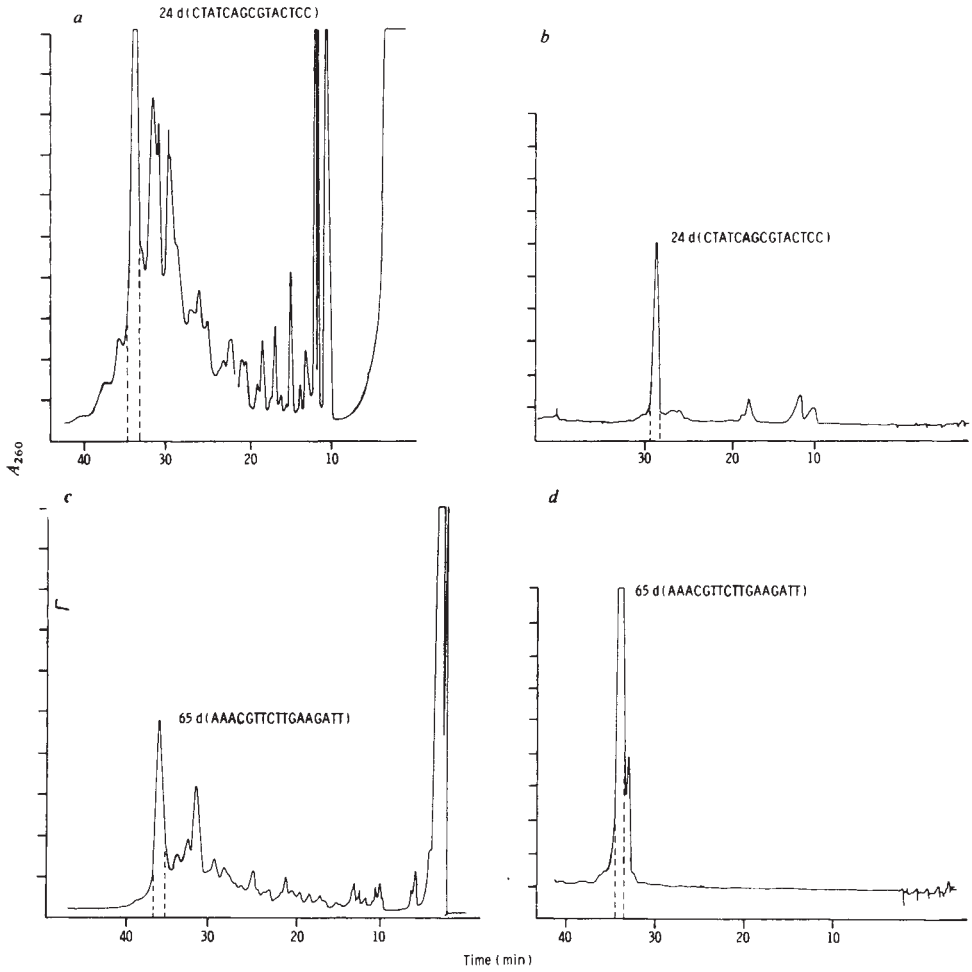
	U		C		A		G		
U	Phe	3 (4)	Ser	3 (4)	Tyr	3 (1)	Cys	3 (4)	U
		5 (4)		4 (4)		1 (3)		2 (1)	C
	Leu	2 (2)		2 (2)	Ochre	1 (1)	Opal	–	A
		0 (4)		0 (0)	Amber	–	Trp	2 (2)	G
C	Leu	1 (0)	Pro	0 (3)	His	2 (2)	Arg	6 (0)	U
		3 (8)		1 (2)		1 (1)		2 (0)	C
		2 (1)		0 (1)	Gln	7 (2)		0 (1)	A
		14 (7)		5 (0)		3 (8)		0 (0)	G
A	Ile	0 (0)	Thr	3 (2)	Asn	2 (2)	Ser	0 (0)	U
		7 (7)		6 (4)		4 (4)		4 (3)	C
		0 (0)		0 (3)	Lys	4 (4)	Arg	3 (6)	A
	Met	6 (6)		0 (0)		4 (4)		1 (5)	G
G	Val	2 (1)	Ala	6 (4)	Asp	4 (5)	Gly	1 (0)	U
		1 (2)		1 (3)		7 (6)		2 (1)	C
		2 (0)		2 (2)	Glu	7 (6)		0 (2)	A
		1 (3)		1 (1)		8 (9)		0 (0)	G

Construction of recombinant plasmids

The synthetic gene contains single-stranded cohesive ends to facilitate ligation to a *Bam*HI–*Sal*I plasmid vector. The plasmid, pPM50, chosen for cloning the synthetic gene, is a derivative of plasmid pAT153 (ref. 40) in which the region of the tetracycline

© 1981 Nature Publishing Group





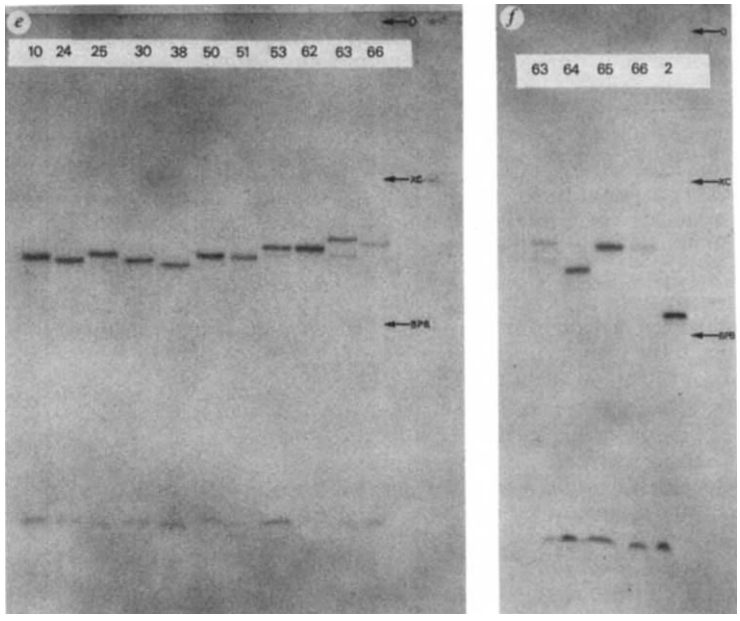


Fig. 2 Synthesis and purification of oligodeoxynucleotides. The 67 oligodeoxynucleotides were synthesized by a solid phase procedure (from a library of dimers) on a polyacrylamide support[33,34]. Functionalized polydimethylacrylamide resin (100 mg), substituted with the appropriate 5′-*O*-dimethoxytrityl-2′-deoxynucleoside via a 3′-*O*-succinamido linkage (~26 μmol total functionality) was swollen in dimethylformamide then treated with phenylisocyanate as described previously[33,34]. Each oligonucleotide assembly involved addition of protected dinucleotide (or mononucleotide) units by the following cycle. The resin was washed sequentially with pyridine (×5); chloroform/methanol (7/3, ×6); 5% benzenesulphonic acid in chloroform/methanol (7/3, for 2×5 s to remove dimethoxytrityl group); dimethylformamide (×4); pyridine (×8). Each washing operation involved a solvent volume of ~6 ml. The resin was dried by co-evaporation with anhydrous pyridine (1 ml) and the dinucleotide anion (100 mg, ~3 equivalents previously dried by pyridine co-evaporation) was added in anhydrous pyridine (4 ml) followed by the coupling agent, mesitylene-3-nitro-1,2,4-triazole[38] (60 mg, ~8 equivalents) or 2,4,6-triisopropylbenzenesulphonyl tetrazole (70 mg, ~8 equivalents). The coupling reaction was allowed to proceed, with shaking, for 1–1.5 h. The cycle was repeated with the appropriate dinucleotide anions until the required sequence was obtained. The resin was finally washed (5 volumes each) with dimethylformamide, methylene chloride and ether, then dried *in vacuo*. (Reactions can be carried out on half the scale described, using 50 mg of functionalized resin and 50 mg of dinucleotide anion.) Each oligonucleotide was cleaved from the solid phase (40 mg of oligomer-substituted resin) and all protecting groups were removed (procedure B, ref. 34) by sequentially reacting with 0.3 M 1,1,4,4-tetramethylguanidinium 4-nitrobenzaldoximate[38] in dioxan/water (1/1, 2 ml) at room temperature for 60 h; concentrated ammonia at 50 °C for 5 h; 80% aqueous acetic acid for 30 min at room temperature. The deprotected oligomer was dissolved in pyridine (0.05 ml), ethanol/water (3/7, 2–4 ml) and suitable aliquots were purified initially by HPLC on Partisil 10-SAX (examples in 2*a* and *c*). Columns were eluted at a flow rate of 2 ml min^{-1} with a gradient from 1 mM to 0.4 M potassium phosphate, *p*H 6.8 in 5% ethanol or *p*H 6.5 in 30% ethanol, over a period of 50 min at ambient temperature. The most retained peak(s) were collected, evaporated, redissolved in water (5 ml) and applied to a μ-Bondapak C_{18} reverse-phase column. The oligonucleotide was further purified by first eluting with 0.1 M ammonium acetate, then with a gradient of 10–25% 0.1 M ammonium acetate/acetonitrile (1/1) over 40 min with a flow rate of 2 ml min^{-1} at ambient temperature (examples in *b* and *d*). After addition of excess triethylamine, oligonucleotide-containing fractions were lyophilized and redissolved in water (1 ml). Aliquots of each solution (80 pmol of oligonucleotide) were evaporated and phosphorylated with 1.1 units of T4-induced polynucleotide kinase in 10 μl of a solution containing 160 pmol [γ-^{32}P]ATP (12.5 Ci $mmol^{-1}$), 0.1 mM spermidine, 20 mM dithiothreitol (DTT), 10 mM $MgCl_2$, 50 mM Tris-HCl (*p*H 9.0) and 0.1 mM EDTA for 45 min at 37 °C. Reactions were terminated by addition of 270 μl of 50 mM Tris-borate (*p*H 8.3), 1 mM EDTA in 80% formamide and heating to 90 °C for 5 min. A 7-μl aliquot was examined on a 20% polyacrylamide gel in 7 M urea (examples in *e* and *f*).

© 1981 Nature Publishing Group

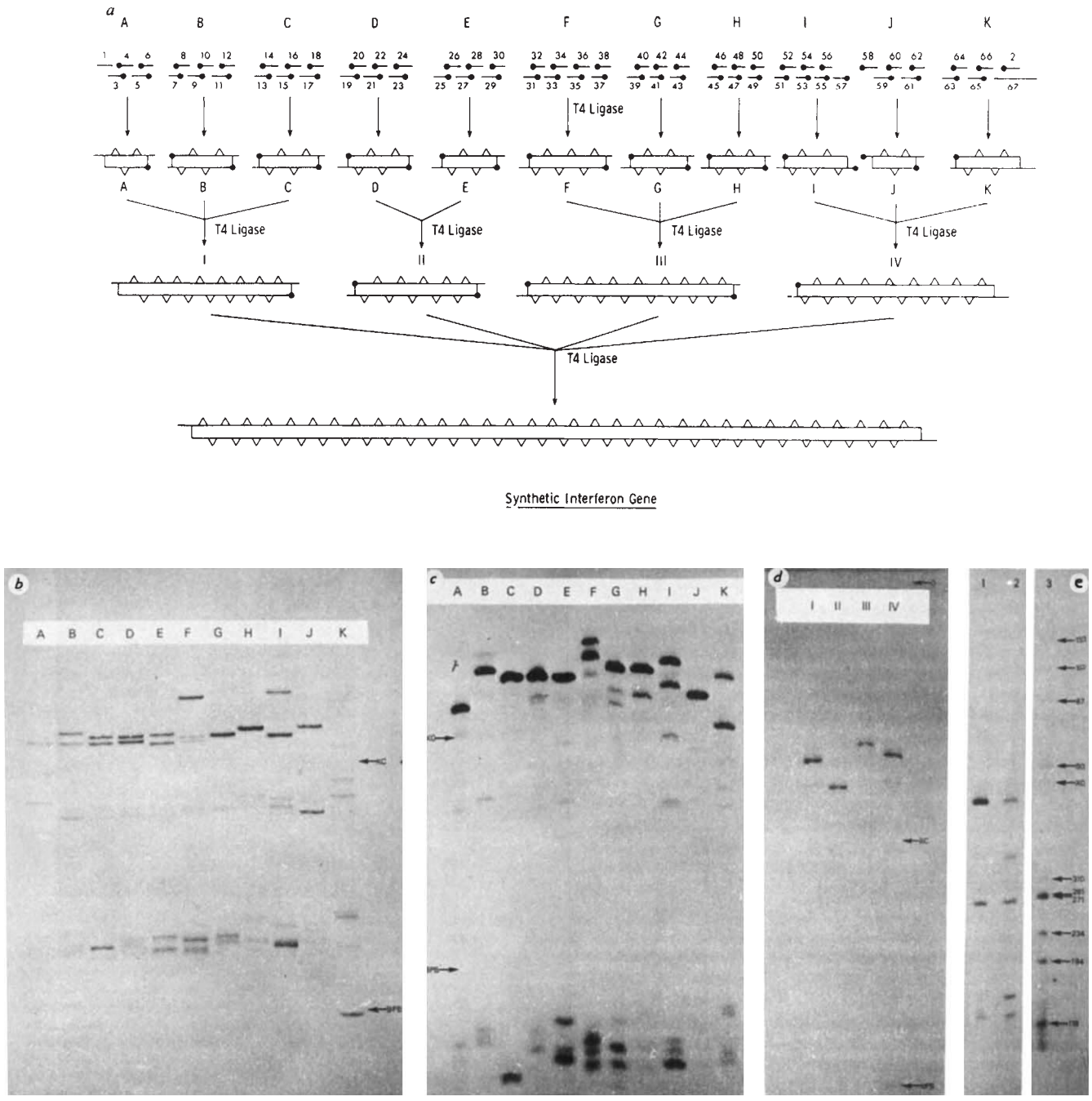


Fig. 3 Ligation of chemically synthesized oligonucleotides. *a*, Ligation strategy. Aliquots of the chemically synthesized oligonucleotides, 2–66 (40 pmol of each), were phosphorylated with 6.5 units of T4-induced polynucleotide kinase in 25 μl of a solution containing 80 pmol [γ-^{32}P]ATP (8.5 Ci mmol^{-1}), 100 μM spermidine, 20 mM DTT, 10 mM $MgCl_2$, 50 mM Tris-HCl (*p*H 9.0) and 0.1 mM EDTA for 30 min at 37 °C. Each oligonucleotide was ethanol precipitated and fractionated by electrophoresis on a 20% polyacrylamide gel in 7 M urea. The ^{32}P-labelled oligomers were sliced from the gel and eluted[46]. Eleven separate T4 ligase-catalysed joining reactions were performed to give fragments A to K. In groups A and K, oligonucleotides 1 and 67, respectively, were not phosphorylated. For the construction of each fragment, the terminal oligonucleotides were present in excess over the internal oligonucleotides. For example, fragment B was constructed by annealing, in 50 μl water, 3.8 pmol of oligonucleotides 7 and 12, 3.6 pmol of 8 and 11 and 3.4 pmol of 9 and 10, heating to 100 °C and cooling slowly to room temperature. Each mixture was dried *in vacuo* and ligated in 20 μl of a solution containing T4 DNA ligase (26 units), 74 mM Tris-HCl (*p*H 7.6), 6.6 mM $MgCl_2$, 10 mM DTT, 2.9 mM ATP, 1.6 mM mercaptoethanol and 1.6 μg bovine serum albumin for 24 h at 22 °C. *b*, Aliquots (1%) of the reaction mixtures were examined on a 15% polyacrylamide gel in 7 M urea. *c*, The DNA from each of the remaining mixtures was ethanol precipitated and fractionated on a 15% polyacrylamide gel in the absence of urea. The DNA from each band was eluted from the gel. An aliquot of each band was re-run in denaturing conditions. In each group, the most slowly migrating band had oligonucleotide strands of the correct size. *d*, Fragments I–IV were constructed from 0.1-pmol of fragments A–K. Annealing was carried out at 45 °C and the mixtures were ligated as described previously. The fragments were isolated by electrophoresis on a 10% polyacrylamide gel in the absence of urea. *e*, The final ligation of fragments I–IV (0.01 pmol of each), to give the complete gene, was carried out as previously described for fragments I–IV, to give the product, at ~500 bp. Lanes 1 and 2 show two separate syntheses of the gene. Lane 3 is a ^{32}P-labelled *Hae*III restriction digest of ΦX174 DNA.

resistance determinant between the *Eco*RI and *Bam*HI sites has been replaced by a 95-bp *Alu*I-generated fragment from the promoter region of the *E. coli* lactose operon. The construction of this plasmid, as shown in Fig. 4*A*, was arranged so that insertion of the *Alu*I fragment resulted in the regeneration of both the *Eco*RI and *Bam*HI recognition sequences. Introduction of the synthetic oligonucleotide between the *Bam*HI and *Sal*I cleavage sites of the pPM50 vector plasmid should therefore place the interferon coding sequence under the transcriptional control of the *lac* promoter in an orientation such that the ATG codon forms a 'hybrid ribosome binding site' with the *lac* promoter Shine–Dalgarno (S–D) sequence[41]. Thus, the

© 1981 Nature Publishing Group

sequence around the *Bam*HI site will be

S–D *Bam*HI Met Cys

5′——AGGAAACAGGATCCATGTGT——3′

A construction of this type should achieve active transcription and translation leading to expression of the interferon coding sequence.

Such promoter fragments containing an *Eco*RI cleavage site in place of the *Bam*HI site have been used to promote expression of the streptomycin resistance gene of the plasmid R300B (P.M., unpublished data). Similarly, plasmid vectors incorporating the analogous *Alu*I fragment from a mutant *lac* promoter (UV5) have been used to promote high level expression of various heterologous genes (see, for instance, refs. 42–44).

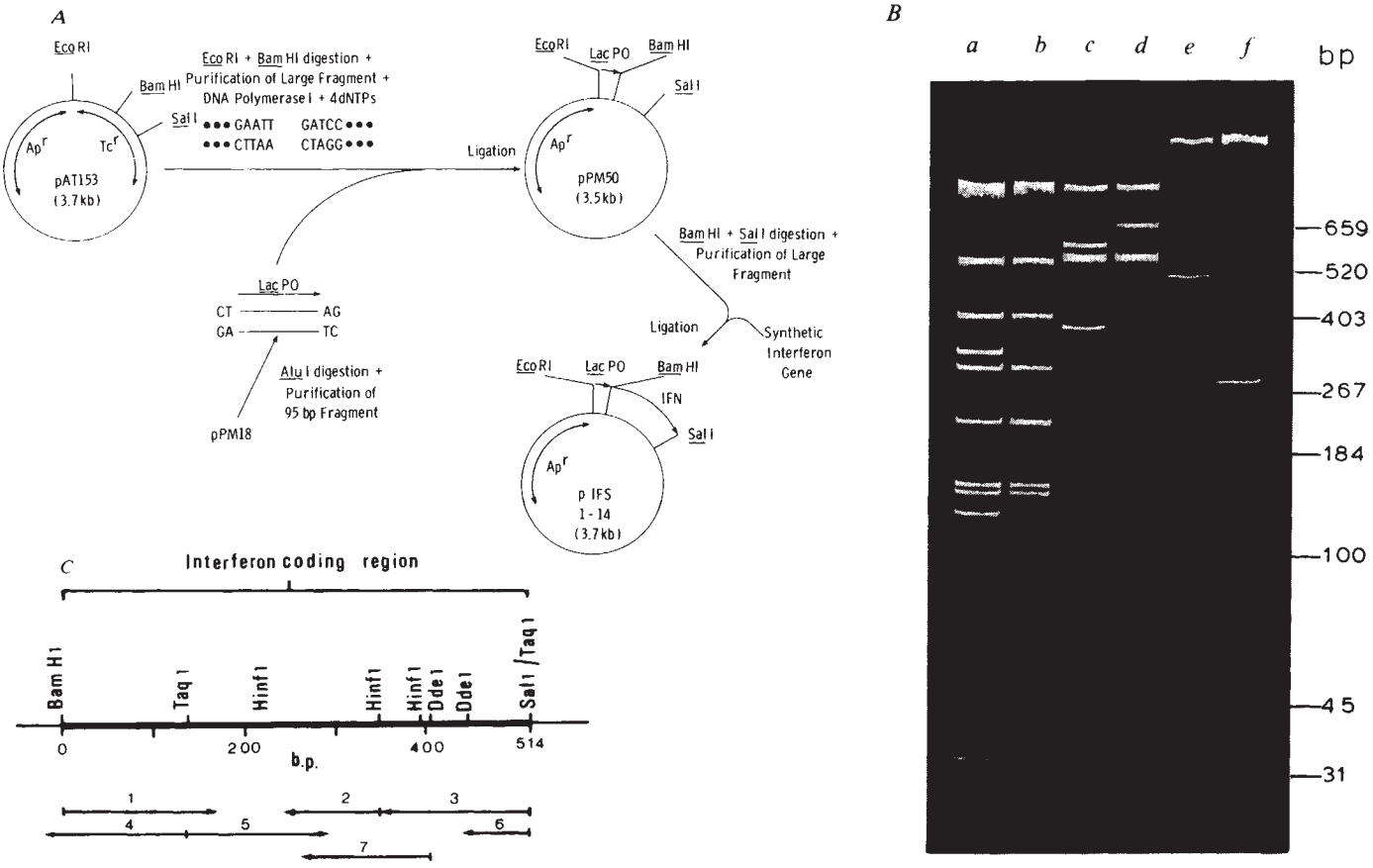


Fig. 4 *A*, Construction of the recombinant pIFS plasmids. DNA (7 μg) of the parental plasmid, pAT153 (ref. 40), was digested sequentially for 60 min at 37 °C with restriction endonucleases *Eco*RI and *Bam*HI in 10 mM Tris-HCl (*p*H 7.6), 6 mM $MgCl_2$ and 1 mM DTT containing 50 mM (*Eco*RI) or 150 mM (*Bam*HI) NaCl. Reactions were terminated by extraction with phenol/chloroform (3/1) and ethanol precipitation. The DNA was resuspended in 30 μl DNA polymerase I buffer (6 mM Tris-HCl (*p*H 7.4), 7 mM $MgCl_2$, 1 mM DTT) containing 40 μM each of dATP, dGTP, dCTP and dTTP to synthesize 3′-oligonucleotide extensions complementary to the protruding 5′-oligonucleotide ends generated by the endonuclease treatment. The reaction was initiated by addition of DNA polymerase I–Klenow fragment (1.5 units) and incubated at 11 °C for 90 min. ^{3}H-dTTP (3 μCi; 16 Ci $mmol^{-1}$) was added so as to follow the extent of reaction by trichloroacetic acid precipitation of 5-μl samples. The reaction was terminated by extraction with phenol/chloroform (3/1) and ethanol precipitation. Plasmid pPM18 is a derivative of the vector plasmid pACYC184 (ref. 49) which has a 6.6-kb *Eco*RI fragment derived from λ*plac*5(ref. 50) inserted at the *Eco*RI site in the plasmid chloramphenicol resistance gene. The λ*plac*5 fragment carries the promoter region of the *E. coli* lactose operon and part of the β-galactosidase gene[51]. Plasmid pPM18 DNA (15 μg) was digested with restriction endonuclease *Alu*I in 6 mM Tris-HCl (*p*H 7.6), 6 mM $MgCl_2$, 50 mM NaCl, 6 mM β-mercaptoethanol and the 95-bp fragment containing the *E. coli* lactose operon promoter/operator region isolated by electroelution[52] from a 5% polyacrylamide gel[53] and purified by chromatography on DE52 DEAE-cellulose[52]. The 'filled in' *Eco*RI/*Bam*HI cleaved pAT153 DNA was blunt-end ligated to the 95-bp *Alu*I fragment in a volume of 50 μl in 20 mM Tris-HCl (*p*H 7.6), 10 mM $MgCl_2$, 10 mM DTT, 5 mM ATP, with T4 DNA ligase (5 units) at 10 °C for 16 h. The ligated mixture was transformed to *E. coli* strain C600 (ref. 54) and ampicillin-resistant transformant colonies selected on antibiotic (25 μg ml^{-1}) minimal X-gal (20 μg ml^{-1}) medium[55]. Colonies containing recombinant plasmids with the *lac* promoter/operator sequence were identified by their blue coloration on the X-gal medium, due to constitutive expression of the chromosomal β-galactosidase gene. Plasmid DNA, isolated from eight independent blue colony transformants, was screened by digestion with restriction endonucleases *Eco*RI and *Bam*HI and one found to contain a single cleavage site for each enzyme separated by ~100 bp. Nucleotide sequence determination of the small (100 bp) fragment generated by digestion with these two enzymes showed that it contained the *lac* promoter/operator sequence flanked by *Eco*RI and *Bam*HI sites and orientated such that transcription would proceed towards the *Bam*HI site (P.M., unpublished data). This plasmid was named pPM50 and was used for the subsequent cloning of the synthetic interferon gene. DNA (4 μg) of plasmid pPM50 was digested with restriction endonucleases *Bam*HI and *Sal*I in 10 mM Tris-HCl (*p*H 7.6), 6 mM $MgCl_2$, 150 mM NaCl and 1 mM DTT at 37 °C for 60 min. The reaction was terminated by extraction with phenol/chloroform and the products subjected to electrophoresis through 1% agarose in 40 mM Tris-HCl (*p*H 7.8), 6 mM sodium acetate, 1 mM EDTA buffer. The larger 3.2-kb fragment was isolated by electroelution[52] and chromatography on DE52 DEAE-cellulose. The purified *Bam*HI–*Sal*I 3.2-kb vector fragment (1 μg) was ligated to the synthetic oligonucleotide preparation with T4 DNA ligase (0.4 units) in a total reaction volume of 30 μl containing 20 mM Tris-HCl (*p*H 7.6), 10 mM $MgCl_2$, 10 mM DTT at 12.5 °C for 16 h. The reaction mixture was used to transform competent cells of *E. coli* K-12, strain MRC 8 (F^-, *dap* 103, *hsd* R, *met* B1, *glm* 533, *upp*1, *dap* 101, *sup* E, *rec* A) and ampicillin-resistance transformants selected on L-agar[55] medium containing *N*-acetylglucosamine (0.02%) and diaminopimelic acid (50 μg ml^{-1}). Transformant colonies containing recombinant plasmids which carried the synthetic oligonucleotide (pIFS plasmids) were identified by Grunstein–Hogness[45] colony hybridization using a radioactively ^{32}P-labelled oligonucleotide from within the synthetic sequence (coding for amino acids 78–97) as hybridization probe. *B*, Structure of plasmids containing the synthetic oligonucleotide. Restriction endonuclease-digested plasmid DNAs were analysed by electrophoresis through 6% acrylamide[53]. *Hinf*I digests of pIFS 1 (*a*) and pPM50 (*b*); *Taq*I digests of pIFS 1 (*c*) and pPM50 (*d*); *Bam*HI and *Sal*I double digests of pIFS 1 (*e*) and pPM50 (*f*). Reaction conditions were as suggested by the enzyme suppliers. *C*, Strategy for sequence determination of the synthetic interferon gene. Top line shows the positions of relevant restriction endonuclease cleavage sites within the synthetic oligonucleotide insert (heavy line) of the recombinant plasmid pIFS1. The nucleotide sequence was determined by use of the Maxam–Gilbert method[46] on fragments labelled at the 3′ end by use of DNA polymerase I (Klenow) with an appropriate [α-^{32}P]-labelled deoxynucleoside triphosphate. The arrows (1–7) indicate the direction and extent of the sequence determined from each labelling position. Fragments sequenced are as follows (the numbers refer to the arrows and the labelled ends of the restriction fragments are designated by an asterisk); 1, **Bam*HI–*Sal*I; 2, *Hinf*I–*Hinf*I*; 3, *Bam*HI–*Sal*I*; 4, *Eco*RI–*Taq*I*; 5, **Taq*I–*Dde*I; 6, *Dde*I–*Taq*I*; 7, *Taq*I–*Dde*I*.

© 1981 Nature Publishing Group

Recombinant plasmids of this type were constructed by ligation of purified *Bam*HI–*Sal*I-cut pPM50 vector DNA to the synthetic oligonucleotide preparation. Two cloning experiments were performed. In one case, plasmid DNA was ligated to a gel-purified 514-bp product of the oligonucleotide ligation mix; in the other, plasmid DNA was mixed with the total oligonucleotide ligation mix and ligation allowed to continue. The ligation mixtures were used for transformation of *E. coli* strain MRC8 and the resulting ampicillin-resistant transformant colonies screened for the presence of the synthetic oligonucleotide insert by Grunstein–Hogness colony hybridization[45] using as hybridization probe a radioactively labelled synthetic oligonucleotide, 60 residues long, coding for amino acids 78–97 of the interferon sequence. Two positively responding colonies were detected among the transformants arising from the ligation of the vector plasmid to the purified 514-bp fragment, and a further 12 from the other ligation. This corresponded to a 10% yield of recombinant plasmid transformants.

Analysis of the plasmid DNA isolated from these colonies showed that all 14 contained equivalent-size plasmids (pIFS 1–14). On digestion with *Bam*HI and *Sal*I endonucleases and examination of the products by gel electrophoresis, all these plasmid DNAs produced an identical 514-bp fragment in addition to the 3.2-kilobase (kb) vector fragment. Four of these plasmids were selected for more detailed analysis (pIFS 1, 2, 4, 14).

Inspection of the theoretical nucleotide sequence of the 514-bp synthetic oligonucleotide (Fig. 1) showed that the d(TCGA) sequence, which is cut by endonuclease *Taq*I, is present twice (at amino acid residues 43/44 and at the *Sal*I restriction site) and the d(GANTC) sequence which is cut by the endonuclease *Hin*fI is present three times (at amino acid residues 72/73, 115/116 and 126/127). Therefore, cleavage of this sequence by these endonucleases should yield an internal 375-bp fragment in the case of *Taq*I and 129-bp and 34-bp internal fragments in the case of *Hin*fI. Examination by acrylamide gel electrophoresis of the DNA fragments produced on treatment of the four selected pIFS plasmid DNAs with these two endonucleases showed that all produced these predicted fragments, whereas the vector plasmid pPM50 produced no such fragments. Figure 4*B* shows these digestion patterns for one of the recombinant plasmids, pIFS 1. We therefore conclude that these four pIFS plasmids must contain a 514-bp insert between the *Bam*HI and *Sal*I cleavage sites which has several widely spaced restriction endonuclease recognition sequences in the positions predicted from the theoretical sequence of the interferon gene oligonucleotide. To confirm that the correct interferon gene sequence was present in these clones, the 514-bp *Bam*HI–*Sal*I fragment from one plasmid, pIFS 1, was subjected to complete nucleotide sequence determination[46] on terminally ^{32}P-labelled DNA fragments according to the strategy shown in Fig. 4*C*. The nucleotide sequence so determined is in complete agreement with the theoretical sequence shown in Fig. 1.

Discussion

The rapid synthesis of genes of this size should allow several interesting problems to be examined. There has been some speculation about the mechanisms by which the expression of human genes cloned into bacteria is controlled (see discussion in ref. 47). Obviously, regulation at the level of transcription or translation initiation will be of fundamental importance and a number of elegant promoter systems have been devised. These have allowed high levels of 'eukaryotic' protein synthesis to be achieved, presumably as a result of the presence of abundant copies of their mRNAs. However, studies by Grantham *et al.*[35–37] comparing the mRNA sequences of highly expressed and weakly expressed genes in bacteria suggest that highly expressed genes have a special strategy for the selection of codons, particulary at position III. This may reflect an optimization of codon–anticodon pairing energies. We have attempted to reproduce this pattern of codon usage in the synthetic gene (Table 1). Others[8] have noted the rarity of the dinucleotide CG in human IFN mRNA, which reflects the strong preference for AGA and AGG as codons for arginine. Whereas the native sequence contains 3 CG residues, 2 of which occur within codons, the synthetic sequence contains 22 CG residues, 14 of which occur within codons. The synthetic sequence also has a marked increase in the incidence of CUG for leucine and of CCG for proline. Insertion of either the natural human DNA sequence or the sequence described herein (or further variations) into the same vector and comparison of expression levels may shed some light on these questions. Examination of, for example, the effect of using a particular codon twice when the same amino acid occurs consecutively in the protein awaits synthesis of further gene variants (see below). It would also be of interest to determine whether any of the complementary sequences within the natural gene (or within the synthetic gene) affects expression. This could involve either formation of internal DNA cruciform structures[48] with resultant influence on transcription or formation of secondary structures at the mRNA level with a resultant influence on translation.

An exciting recent development concerning interferon has been the construction of hybrid genes. Cleavage at a convenient common *Bgl*II site within the coding sequences of a number of the IFN-α genes[23,24] allows fusion of the N-terminal sequence from one gene with the C-terminal sequence from another. Thus, in addition to the range of natural IFN-α molecules with their different pharmacological profiles[11], a potentially large number of hybrid interferons with further distinct properties are accessible. The availability of a large pool of synthetic fragments clearly extends the range of interferon analogues which can be prepared. It should be possible to compare rapidly the relative activities of proteins generated by this approach in a common vector system and it will be interesting to determine whether the natural interferons can be improved upon. The wider implication is that classical medicinal chemistry structure–activity analysis should be possible in relatively large peptides. This may yield considerable benefits in a number of important pharmaceutical areas.

We thank Dr T. F. W. McKillop for helpful discussions and for his continuous support, Drs E. C. Dart, J. S. Morley, W. J. Brammar, J. D. Windass and K. Atherton for their interest and support, Dr D. J. Gilman for computer analyses, and Dr S. Brenner for *E. coli* K-12.

Received 8 June; accepted 14 July 1981.

1. Isaacs, A. & Lindenmann, J. *Proc. R. Soc.* **B147**, 258–267 (1957).
2. Stewart, W. E. II *The Interferon System* (Springer, New York, 1979).
3. Gresser, I. & Tovey, M. G. *Biochim. biophys. Acta* **516**, 231–247 (1978).
4. Bloom, B. R. *Nature* **284**, 593–595 (1980).
5. Taniguchi, T. *et al. Proc. Japan Acad.* **B55**, 464–469 (1979).
6. Taniguchi, T., Ohno, S., Fujii-Kuriyama, Y. & Muramatsu, H. *Gene* **10**, 11–15 (1980).
7. Mantei, N. *et al. Gene* **10**, 1–10 (1980).
8. Derynck, R. *et al. Nature* **285**, 542–547 (1980).
9. Taniguchi, T. *et al. Nature* **285**, 547–549 (1980).
10. Weissenbach, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7152–7156 (1980).
11. Streuli, M., Nagata, S. & Weissmann, C. *Science* **209**, 1343–1347 (1980).
12. Houghton, M. *et al. Nucleic Acids Res.* **8**, 1913–1931 (1980).
13. Houghton, M. *et al. Nucleic Acids Res.* **8**, 2885–2894 (1980).
14. Goeddel, D. V. *et al. Nature* **287**, 411–416 (1980).
15. Nagata, S. *et al. Nature* **284**, 316–320 (1980).
16. Derynck, R. *et al. Nature* **287**, 193–197 (1980).
17. Goeddel, D. V., Shepard, H. M., Yelverton, E., Leung, D. & Crea, R. *Nucleic Acids Res.* **8**, 4057–4074 (1980).
18. Taniguchi, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 5230–5233 (1980).
19. Yelverton, E., Leung, D., Weck, P., Gray, P. W. & Goeddel, D. V. *Nucleic Acids Res.* **9**, 731–741 (1981).
20. Stewart, W. E. II *et al. Gene* **11**, 181–186 (1980).
21. Masucci, M. G. *et al. Science* **209**, 1431–1435 (1980).
22. Allen, G. & Fantes, K. H. *Nature* **287**, 408–411 (1980).
23. Nagata, S., Mantei, N. & Weissmann, C. *Nature* **287**, 401–408 (1980).
24. Goeddel, D. V. *et al. Nature* **290**, 20–26 (1980).
25. Houghton, M. *et al. Nucleic Acids Res.* **9**, 247–266 (1981).
26. Tavernier, J., Derynck, R. & Fiers, W. *Nucleic Acids Res.* **9**, 461–471 (1981).
27. Knight, E., Hunkapiller, M. W., Korant, B. D., Hardy, R. W. F. & Hood, L. E. *Science* **207**, 525–526 (1980).
28. Zoon, K. C. *et al. Science* **207**, 527–528 (1980).
29. Itakura, K. *et al. Science* **198**, 1056–1063 (1977).
30. Goeddel, D. V. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 106–110 (1979).
31. Wetzel, R. *et al. Biochemistry* **19**, 6096–6104 (1980).
32. Itakura, K. & Riggs, A. D. *Science* **209**, 1401–1405 (1980).
33. Markham, A. F. *et al. Nucleic Acids Res.* **8**, 5193–5205 (1980).
34. Gait, M. J. *et al. Nucleic Acids Res.* **8**, 1081–1096 (1980).
35. Grantham, R., Gautier, C., Gouy, M., Mercier, R. & Pavé, A. *Nucleic Acids Res.* **8**, r49–r62 (1980).
36. Grantham, R., Gautier, C. & Gouy, M. *Nucleic Acids Res.* **8**, 1893–1912 (1980).

© 1981 Nature Publishing Group

37. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. *Nucleic Acids Res.* **9**, r43–r74 (1981).
38. Reese, C. B., Titmas, R. C. & Yau, L. *Tetrahedron Lett.*, 2727–2730 (1978).
39. Frank, R. & Köster, H. *Nucleic Acids Res.* **6**, 2069–2087 (1979).
40. Twigg, A. J. & Sherratt, D. *Nature* **283**, 216–218 (1980).
41. Shine, J. & Dalgarno, L. *Nature* **254**, 34–38 (1975).
42. Backman, K., Ptashne, M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4174–4178 (1976).
43. Roberts, T. M., Kacich, R. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 760–764 (1979).
44. Goeddel, D. V. *et al. Nature* **281**, 544–548 (1979).
45. Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
46. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1978).
47. Morgan, J. & Whelan, W. J. *Recombinant DNA and Genetic Experimentation*, 141–143 (Pergamon, Oxford, 1979).
48. Panayotatis, N. & Wells, R. D. *Nature* **289**, 466–470 (1981).
49. Chang, A. C. Y. & Cohen, S. N. *J. Bact.* **134**, 1141–1156 (1978).
50. Shapiro, J. *et al. Nature* **224**, 768–774 (1969).
51. Helling, R. B., Goodman, H. M. & Boyer, H. W. *J. Virol.* **14**, 1235–1244 (1974).
52. Smith, H. O. *Meth. Enzym.* **65**, 371–380 (1980).
53. Maniatis, T., Jeffrey, A. & Van deSande, H. *Biochemistry* **14**, 3787–3794 (1975).
54. Bachmann, B. J. *Bact. Rev.* **36**, 525–557 (1972).
55. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor, New York, 1972).

LETTERS TO NATURE

Regular patchy distribution of cytochrome oxidase staining in primary visual cortex of macaque monkey

Jonathan C. Horton & David H. Hubel

Department of Neurobiology, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA

In two preliminary studies, normal macaque[1] and squirrel-monkey[2,3] striate cortex cut parallel to the surface and stained for cytochrome oxidase (a mitochondrial enzyme) showed a striking pattern of regularly spaced patches. This was surprising, since until then no physiological or anatomical studies had suggested such a patchy organization. In the present study in the macaque we found that the patches were arranged in rows spaced about 350 μm apart. When one eye was injected with tritiated proline the rows of patches in layers II and III lay in register with the ocular-dominance bands seen autoradiographically in layer IVc. Removing one eye caused the patches in every other row to shrink and blanch. The rows of patches are therefore centred on the ocular dominance columns. Regions labelled by 2-deoxyglucose autoradiography after stimulating one eye with black-and-white stripes in all orientations consisted of rows of patches that lay in register with the cytochrome oxidase patches in every other row. On stimulating monkeys with stripes of a single orientation, the deoxyglucose-labelled regions formed a lattice that included the cytochrome oxidase patches but was more extensive. Thus either the deoxyglucose is not labelling the orientation columns at all, or the orientation columns coalesce in the areas marked by the cytochrome oxidase stain.

Wong-Riley has used a stain for cytochrome oxidase to demonstrate a pattern of alternating light and dark bands in layer IV of a monocularly deprived kitten, the lighter bands presumably reflecting a lowered metabolic activity in the set of columns belonging to the closed eye[4]. Our original purpose was to examine cytochrome oxidase activity in the different layers of the striate cortex of the normal macaque monkey. Sections (50 μm) from glutaraldehyde–paraformaldehyde fixed tissue were cut on a freezing microtome and processed for cytochrome oxidase activity following a procedure modified from Seligman *et al.*[5]. In sections cut perpendicular to the cortical surface, cytochrome oxidase staining was darkest in layers IVa and IVc, which receive the major direct projections from the lateral geniculate body[6]. Layers II and III were lightly stained but showed periodic fluctuations in density. To examine the pattern more closely we sectioned the cortex tangentially. Cytochrome oxidase staining showed an array of dark oval patches about 150 × 200 μm which were most obvious in Layers II and III but were also present, though faint, in layer VI. Although varying to some extent from animal to animal, the patches were generally aligned in rows spaced ~350 μm apart, with the long axis of the patches parallel to the rows. Within a row the patches were ~550 μm apart, but in some places became confluent. Occasionally, patches in neighbouring rows seemed to be aligned, forming a square or hexagonal array over small regions of cortex.

The rows of cytochrome oxidase patches resembled ocular dominance columns in their spacing and in intersecting the 17–18 border at right angles (Fig 1*a*). It was obviously important to learn whether these rows were really related to the ocular dominance columns, and, if so, whether the patches lay along the dominance columns or straddled the borders between them. We therefore removed one eye of a macaque monkey, killed the animal 10 days later and stained the visual cortex for cytochrome oxidase. Sections through layer IVc showed ocular dominance columns visable as alternating light and dark bands (Fig. 1*b*), in striking contrast to the uniformly deep staining of this layer for cytochrome oxidase in normal monkeys. In sections passing through layers II and III there were parallel

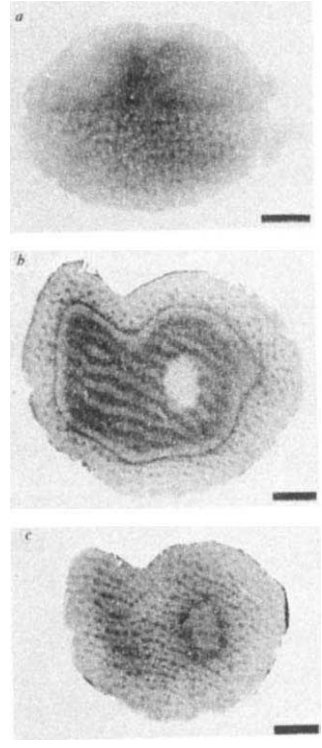


Fig. 1 *a*, Tangential section through the superficial layers of the visual cortex of the normal macaque monkey; cytochrome oxidase stain. The section passes through the 17–18 border, which runs horizontally in the figure with area 17 below and 18 above. The patches in 17 are aligned in rows perpendicular to the 17–18 border. *b*, Cytochrome oxidase section tangential to area 17 in a macaque monkey that had its right eye removed 10 days previously. Section grazes layer V, which forms a pale oval near the centre, surrounded from within out by layers IVc (dark stripes), IVb (pale and narrow), IVa (dark, very narrow) and II, III (containing patches). Dark and pale alternating stripes in layer IVc are typical of ocular dominance columns, in their regularity, direction and spacing. *c*, Same monkey as in *b*. This section is more superficial, grazing layer IVb. Cytochrome oxidase patches lie in rows that are alternately dark and pale: the pale rows lie over the light stripes in layer IVc corresponding to the injured eye. Scale bar, 2 mm.

0028–0836/81/340762—08$01.00

© 1981 Macmillan Journals Ltd

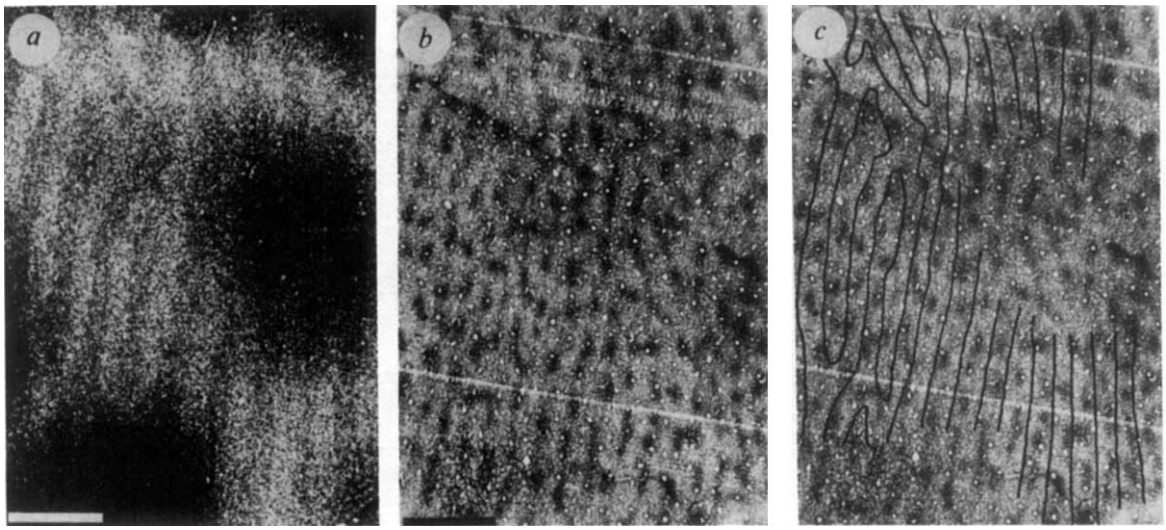


Fig. 2 *a*, Dark-field autoradiograph of striate cortex in a normal monkey whose right eye was injected with ^{3}H-proline 2 weeks earlier. The section is tangential, grazing layer V (dark ovals) but passing mainly through layer IVc. Here typical ocular dominance columns are seen as light bands of label corresponding to the injected eye separated by darker gaps. *b*, More superficial section from the same block as *a*, cytochrome oxidase stain. The rows of cytochrome oxidase patches in the superficial layers follow the pattern of the ocular dominance patches. This is shown in *c* by drawing the borders of the columns from *a* directly onto *b*. Scale bar, 1 mm.

rows of patches as in normal monkeys, but in every other row the patches were paler and smaller (Fig. 1*c*). When one of these sections was aligned with a section through layer IVc, using as a guide small blood vessels (which generally run normal to the cortex through its full thickness), the alternating rows of dark and light patches fell into precise register with the dark and light bands in layer IVc. This indicated that the rows of patches in normal monkeys lie centred on the ocular dominance columns rather than along the borders separating them. All the cells in any given patch are thus likely to be strongly dominated by the same eye.

In a second monkey we injected one eye with 2 mCi ^{3}H-proline to label the ocular dominance columns in layer IVc by transneuronal autoradiography. Figure 2*a*, a dark-field autoradiograph of a section tangential to layer IVc, shows typical ocular dominance columns in which the light bands represent the injected eye. A more superficial section stained for cytochrome oxidase shows the expected array of patches (Fig. 2*b*). When aligned with the autoradiograph, again using radial blood vessels as a guide, the rows of patches follow the course of the ocular dominance columns and are centred on them (Fig. 2*c*).

In a third monkey we used the 2-deoxyglucose method to label regions of cortex activated by stimulation of one eye. We anaesthetized the animal with sodium thiopental injected ^{14}C-2-deoxyglucose (100 μCi per kg), and stimulated the visual field of the right eye with a set of black-and-white stripes of irregular width and spacing, moved slowly back and forth and steadily rotated so as to expose the animal to all orientations about once every minute. After 45 min the monkey was killed, the brain perfused, frozen and sectioned and the dried sections pressed against X-ray film. Alternate sections were stained for cytochrome oxidase. A pattern representing ocular dominance columns was visible in the deoxyglucose autoradiographs of tangential sections through layer IVc. An adjacent section stained for cytochrome oxidase showed no trace of ocular dominance columns in layer IVc, indicating that the brief period of monocular stimulation was insufficient to affect cytochrome oxidase levels. When a cytochrome oxidase section through layers II and III (Fig. 3*a*) was superimposed on a deoxyglucose autoradiographs of layer IVc, the rows of cytochrome oxidase patches were again centred over the ocular dominance columns, confirming the results of the eye removal and transneuronal autoradiography experiments.

Deoxyglucose sections adjacent to Fig. 3*a* showed an array of patches of increased deoxyglucose uptake (Fig. 3*b*) which lay

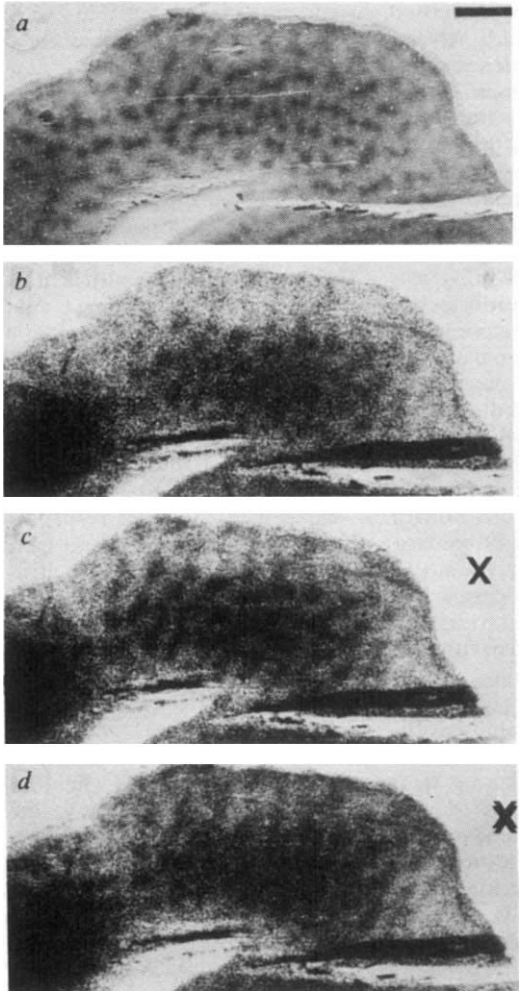


Fig. 3 *a*, An array of cytochrome oxidase patches in layers II, III of striate cortex of a normal macaque monkey. *b*, This 2-deoxyglucose autoradiograph of an adjacent section shows patches of uptake of label after stimulation of one eye with black-and-white stripes in all orientations. *c*, When *a* and *b*, photographed on film, are overlapped in precise register the two patterns reinforce. *d*. When shifted out of alignment by 150 μm (see Xs) the patterns tend to cancel. Scale bar, 1 mm.

© 1981 Nature Publishing Group

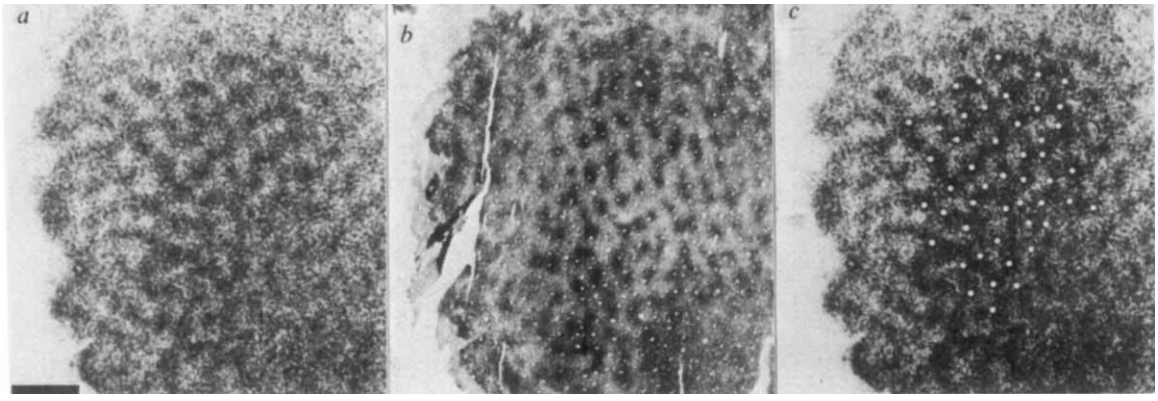


Fig. 4 *a*, Complex pattern of ^{14}C-2-deoxyglucose label in layers II, III of macaque striate cortex after stimulation of both eyes with vertical stripes. *b*, An adjacent section stained for cytochrome oxidase shows an array of patches. *c*, The cytochrome oxidase patches fall within the lattice of deoxyglucose label, as shown by representing each patch in *b*, with a small white dot, and placing them directly on *a*. Scale bar, 1 mm.

over the deoxyglucose-labelled ocular dominance columns in layer IVc. A similar patchy pattern from stimulating one eye with all orientations has been observed by Kennedy *et al.*[7] and by Hendrickson and Wilson[8]. The deoxyglucose pattern resembled the cytochrome oxidase pattern, but the patches were larger and the rows more widely spaced. When Figs 3*a* and 3*b* were superimposed, in and out of register (Fig. 3*c*, *d*), the patches of increased deoxyglucose label matched the cytochrome patches lying in every other row.

Attempts to learn whether there was any relationship between the cytochrome oxidase patches and orientation columns produced unexpected results. In two experiments we stimulated with vertical black-and-white stripes after injecting 2-deoxyglucose. As shown in Fig 4*a*, the pattern in tangential sections through layers II and III was periodic and yet highly complex, forming a lattice of stripes, rosettes and patches, as reported previously[9]. The patches seen in adjacent sections stained for cytochrome oxidase (Fig. 4*b*) fell along the lattice of deoxyglucose label (Fig. 4*c*). Where the deoxyglucose pattern appeared patchy, the patches tended to coincide with the cytochrome oxidase patches; the deoxyglucose pattern thus included the cytochrome patches but seemed to extend beyond and between them. We first thought that the cytochrome oxidase patches might have some special relationship to vertical orientation columns[2], but two similar experiments using deoxyglucose and horizontal stripes gave the same result, with the cytochrome oxidase patches lying along the lattice formed by the deoxyglucose label. Thus all orientations of line stimuli probably activate cortex in the regions of the patches, as indeed is suggested by the deoxyglucose experiment in which one eye was stimulated with all orientations. A combination of physiological recordings and a double-label deoxyglucose technique[10] will probably help clarify the relationship between the patches and the orientation columns.

In squirrel monkeys Humphrey and Hendrickson[3] have shown that the deoxyglucose pattern obtained when both eyes are stimulated with all orientations is identical to the cytochrome oxidase pattern. Similarly, we have found in two squirrel monkeys a strong correlation between the cytochrome oxidase patches and the pattern after stimulation with horizontal stripes in both eyes. In the human and the galago, a prosimian primate, the striate cortex likewise shows cytochrome oxidase patches, we find that the cat and tree shrew lack such patches. A patchy distribution of cytochrome oxidase staining may reflect a system of inputs, outputs and intrinsic connections unique to the visual cortex of primates.

We thank Birthe Storai, Suzanne Fenstemaker and Debra Hamburger for technical help, Marc Peloquin for photography, Olivia Brum for typing the manuscript, and Helen Sherk, Simon LeVay and Torsten Wiesel for advice and criticism. This study was funded by NIH grant EY 00605; J.C.H. is supported by Medical Scientist Training Program grant GM 02220.

Received 6 April; accepted 6 July 1981.

1. Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **160**, 106–154 (1962), **165**, 559–578 (1963), **195**, 215–243 (1968); *J. Comp. Neurol.* **158**, 267–294 (1974).
2. Horton, J. C. & Hubel, D. H. *Soc. Neurosci.* **113**, 5 (1980).
3. Humphrey, A. L. & Hendrickson, A. E. *Soc. Neurosci.* **113**, 6 (1980).
4. Wong-Riley, M. *Brain Res.* **171**, 11–28 (1979).
5. Seligman, A. N. *et al. J. Cell Biol.* **38**, 1–14 (1968).
6. Hubel, D. H. & Wiesel, T. N. *J. comp. Neurol.* **146**, 421–450 (1972).
7. Kennedy, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **73**, 4230–4234 (1976).
8. Hendrickson, A. E. & Wilson, J. R. *Brain Res.* **170**, 353–358 (1979).
9. Hubel, D. H., Wiesel, T. N. & Stryker, M. P. *J. comp. Neurol.* **177**, 361–380 (1978).
10. Livingstone, M. S. & Hubel, D. H. *Nature* **291**, 554–561 (1981).

Effect of chemical environment on levels of substance P and somatostatin in cultured sensory neurones

Anne W. Mudge*

Departments of Physiology and Pharmacology, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA

The neurotransmitter produced by autonomic neurones is influenced by the environment in which they develop[1]. For example, Le Douarin[2] and her colleagues have shown that regions of the neural crest which normally give rise to cholinergic ganglia can be induced to develop into adrenergic cells by transplanting them into a region of the crest which normally gives rise to adrenergic ganglia. Similarly, sympathetic neurones grown in culture express either adrenergic or cholinergic properties depending on whether they are grown in the absence or presence of certain types of non-neuronal cells[1,3–5]. Patterson *et al.*[6,7] demonstrated that this is not due to selective survival of a population of neurones but that the transmitter choice of individual neurones can be altered by soluble factors produced by non-neuronal cells. Like autonomic neurones, sensory neurones derive from the neural crest and it therefore seems likely that the type of transmitter they produce could also be influenced by the environment in which they develop. Here I demonstrate that when sensory neurones from embryonic chick dorsal root ganglia are grown together with ganglionic non-neuronal cells or with medium 'conditioned' by incubation with such cells, they produce increased amounts of somatostatin (SOM). This increase is neither accompanied by an increase in substance P (SP) content nor a detectable change in neuronal survival and thus differs from the effect of nerve growth factor (NGF), which increases survival of sensory neurones without affecting the relative levels of SOM and SP.

* Present address: MRC Neuroimmunology Project, Department of Zoology, University College London, Gower Street, London WC1E 6BT, UK.

0028–0836/81/340764—04$01.00

© 1981 Macmillan Journals Ltd

Although the transmitters used by sensory neurones are unknown, several peptides, namely substance P, somatostatin, cholecystokinin (CCK) and vasoactive intestinal peptide (VIP), are present in small-diameter neurones of the DRG (dorsal root ganglia) and in small-diameter fibres that project into the spinal cord dorsal horn[8]. Although SP and SOM seem to be in separate populations of neurones[9], it is not known which subclasses of small fibres contain these peptides or whether they coexist with other transmitters. As both SP and SOM can be released from sensory neurones[10–13] and both can alter neuronal firing rates in the dorsal horn[14,15], it seems likely that they (and probably also CCK and VIP[16]) are involved in synaptic transmission in the spinal cord, although their functions remain unclear.

DRG from 10-day-old chicken embryos were dissociated to single cells and plated on to collagen-coated tissue culture dishes

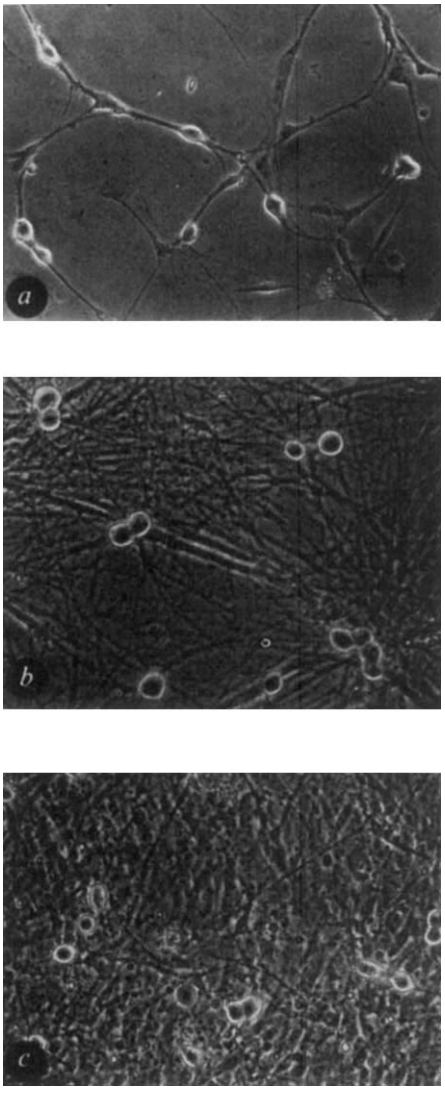


Fig. 1 Phase contrast micrographs of cultures of chick DRG. Dissociated cells were plated on to collagen-coated 35-mm culture dishes (Falcon). The growth medium used was Eagle's minimum essential medium (MEM) supplemented with 2 mM glutamine, 50 μg ml^{-1} penicillin, 50 U ml^{-1} streptomycin, 10% horse serum (all Gibco), 5% chick embryo extract (CEE) plus nerve growth factor (NGF). *a*, Neurones plus ganglionic non-neuronal cells 24 h after plating. *b*, Neurones plus ganglionic non-neuronal cells 15 days after plating. The dividing non-neuronal cells were confluent 6 days after plating, at which time the cultures were irradiated with ^{60}Co to arrest further cell division. *c*, Neurones-alone cultures 15 days after plating. These cultures were treated with 5×10^{-6} M cytosine arabinoside for the period between 12 h and 5 days after plating to kill almost all the non-neuronal cells. Cultures from the same plating as in *a*, *b* and *c* provided the data shown in Fig. 2. Scale bar, 50 μm.

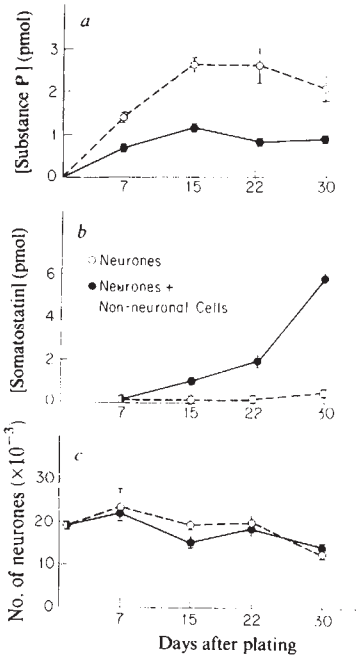


Fig. 2 SP (*a*) and SOM (*b*) content of DRG cultures grown for the number of days indicated in either the presence (●) or virtual absence (○) of ganglionic non-neuronal cells. The number of neurones present in each culture is shown in *c*. Data are expressed as mean ±s.e.m. of either three or four cultures; all cultures were from the same plating.

to give two types of culture: (1) cultures of 'neurones alone' in which almost all the non-neuronal cells were eliminated by treatment with cytosine arabinoside; (2) neurones grown together with dividing ganglionic non-neuronal cells. Within 24 h after plating, both neurones and non-neuronal cells had attached to the collagen substratum and phase-bright, neurone-like cells had extended processes (Fig. 1*a*) which grew during the next 2 weeks to form a dense network (Fig. 1*b*, *c*).

At various times after plating, the number of phase-bright neuronal somata was estimated by counting about 30 fields (×200 magnification) chosen at random and normalizing for the total area of the dish. In typical platings there were 15–20 neurones per field and these were evenly distributed throughout the dish. The cultures were then extracted using 2M acetic acid and the extracts lyophilized. The amount of SP and SOM present in a culture was assayed by radioimmunoassay. Antibody R6P was used for the SP assay as previously described[17]. The SOM assay used was a modification of that described elsewhere[18]. Briefly, the SOM assay used antibody M4 with ^{125}I-tyrosinyl–SOM as a tracer. The incubation and separation procedures were essentially the same as those used for the SP assay. SOM standards were prepared using cyclic-SOM (Boehringer Mannheim) and the concentrations checked by amino acid analysis. When extracts of neurones plus non-neuronal cell cultures were chromatographed on a gel-filtration column (Sephadex G-25-80; Sigma), over 85% of both the SP- and SOM-immunoreactive material eluted as single peaks in the same volume as either synthetic SP or synthetic SOM. (The fact that only one peak of SOM is present in the extracts of DRG cultures is noteworthy, as a variety of tissues, including adrenal medulla[19], contain SOM-immunoreactive material of several different molecular weights when assaved using the same antibody).

The peptide content of cultures containing either neurones alone or neurones plus non-neuronal cells is shown in Fig. 2. The cells plated on to each dish (from five ganglia) contained 35 fmol of SP and 120 fmol of SOM when freshly dissociated. During the next 2 weeks, the content of SP in the cultures of neurones alone increased ~75-fold and then reached a plateau, while the content of SOM remained low; the ratio of SOM/SP was 0.03 at

© 1981 Nature Publishing Group

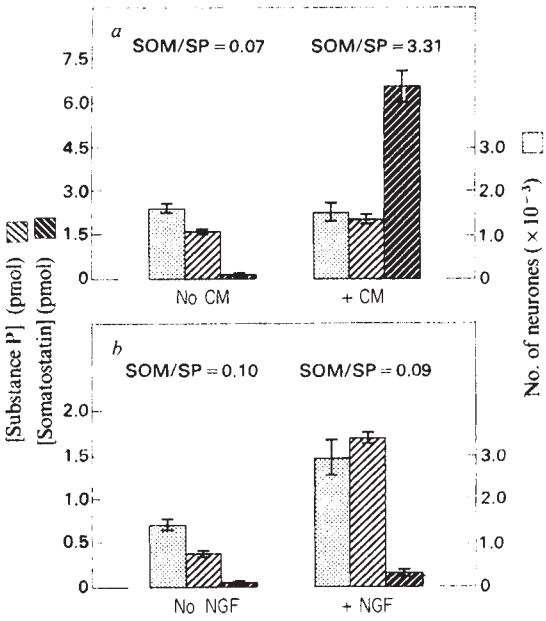


Fig. 3 Effect of conditioned medium and NGF on levels of SP and SOM and on DRG neuronal survival. *a*, Neurones-alone cultures were fed every day for 11 days with medium containing 33% conditioned medium (CM); control cultures received only fresh medium (all media contained 1 $\mu g\ ml^{-1}$ NGF). Results are mean ±s.e.m. of triplicate cultures. Ganglionic non-neuronal cells without neurones were obtained by removing from the cultures neurones, together with non-neuronal cells, using trypsin, filtering the mixture through lens paper, re-plating and irradiating the cells when confluent; neurones do not survive this treatment. Medium incubated for 2 days with such cultures provided the conditioned medium. Aliquots of conditioned medium and the ganglionic non-neuronal cells were also assayed for SP and SOM. *b*, Neurones-alone cultures were fed for 11 days with medium either with or without the addition of NGF at a protein concentration of 1 $\mu g\ ml^{-1}$. NGF was prepared from mouse salivary glands[26]. Results are mean ±s.e.m. of triplicate cultures.

2 weeks and 0.13 at 4 weeks. However, when the neurones were grown together with ganglionic non-neuronal cells, the content of SOM steadily increased after the first week of culture (the time at which the non-neuronal cells are confluent) and after 2 weeks and 4 weeks, the ratio of SOM/SP was 0.96 and 6.34, respectively (Fig. 2*a*, *b*). This increase in SOM content could not be attributed to a general effect on neuronal growth, because it was not associated with an increase in SP content. Indeed, the content of SP in the presence of non-neuronal cells was suppressed by ~50% at all time points (Fig. 2*a*). The number of neurones surviving in culture remained constant from the first day after plating, for 3 weeks in culture, while in the fourth week there was a 20% loss of neurones (Fig. 2*c*). There was no detectable difference within the limits of the counting procedure in neuronal survival in the presence or absence of non-neuronal cells.

To determine whether the non-neuronal cells had to be in contact with the neurones to influence the production of SOM, cultures of neurones alone were fed with medium that had been conditioned by incubation with ganglionic non-neuronal cells without neurones. Such conditioned medium markedly increased the SOM/SP ratio without detectably influencing neuronal survival (Fig. 3*a*). This experiment also showed that it was the neurones which increased their SOM content, as neither the non-neuronal cells nor the conditioned medium contained detectable amounts of SOM or SP.

Although NGF is known to affect the development[20,21] and SP content[22,23] of sensory neurones and is produced by cultured non-neuronal cells[20], it is unlikely to be the agent responsible for the change in relative amounts of SOM and SP in these experiments. When NGF was added to cultures of neurones alone, it doubled the number of neurones which survived and increased both SOM and SP proportionately, but did not significantly alter the ratio of SOM/SP (Fig. 3*b*). Moreover, all the experiments demonstrating the effect of non-neuronal cells or conditioned medium on the SOM/SP ratio were done in saturating levels of NGF.

The survival of sensory neurones depends on additional factors distinct from NGF which are present in brain extracts and secreted by a glioma cell line[21]. In my experiments, the 5% chick embryo extract (CEE) present in the medium presumably provided some factor which supported the survival of neurones in the absence of added NGF. As CEE was present in all the experiments reported here and, again, because the change in SOM/SP ratio occurred without a change in neuronal survival, the 'survival factor' demonstrated by Barde *et al.*[21] is also unlikely to be the factor which increased SOM content. Furthermore, when an extract of embryonic chick brain[24] (100 $\mu g\ ml^{-1}$ protein) was added to cultures of neurones alone for 2–10 days after plating, SP content was increased ~two-fold and SOM rather less; thus there was little effect on the SOM/SP ratio.

These results demonstrate that, like sympathetic neurones, sensory neurones respond to a factor from non-neuronal cells that dramatically influences peptide synthesis. This change in the relative content of SOM and SP is distinct from previously described influences on neuronal survival in the ganglia. It is not known whether, analogous to the switch in adrenergic–cholinergic properties of sympathetic neurones[1], a single population of developing sensory neurones is capable of making either SP or SOM. It is also unknown whether the levels of CCK and VIP can be similarly influenced, nor is it clear what part this effect plays in normal development of sensory neurones. However, in view of the diversity of peptides present in sensory (and autonomic) ganglia, it is possible that the soluble factor described here may represent one of the differentiation signals required for development of neural crest cells.

As DRG non-neuronal cells consist of two major cell types, glial cells and fibroblasts, it is of interest to know which type of cell is influencing the relative levels of SOM and SP. Experiments using purified cultures of either Schwann cells or fibroblasts from DRG[25] or autonomic ganglia should establish this point and provide a well defined system for extending these observations.

I thank Drs G. D. Fischbach and S. E. Leeman for their support and encouragement of this work. I also thank P. H. Patterson for his encouragement and advice. The somatostatin antibody was a gift from M. Arnold and J. Fernstrom. This work was supported by NIH grant NS11160 to G. D. Fischbach and AM1651007 to S. E. Leeman. A.W.M. was supported by NIH training grants GM00919 and NS07009.

Received 6 May; accepted 8 July 1981.

1. Patterson, P. H. *A. Rev. Neurosci.* **1**, 1–17 (1978).
2. Le Douarin, N. M. *Nature* **286**, 663–669 (1980).
3. Varon, S. S. & Bunge, R. P. *A. Rev. Neurosci.* **1**, 327–361 (1978).
4. Furshpan, E. J., Macleish, P. R., O'Lague, P. H. & Potter, D. D. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4225–4229 (1976).
5. Landis, S. C. *Devl Biol.* **77**, 349–361 (1980).
6. Patterson, P. H. & Chun, L. L. Y. *Devl Biol.* **56**, 263–280 (1977).
7. Reichardt, L. F. & Patterson, P. H. *Nature* **270**, 147–151 (1977).
8. Hokfelt, T., Johansson, O., Ljungdahl, A., Lundberg, J. M. & Schultzberg, M. *Nature* **284**, 515 (1980).
9. Hokfelt, T. *et al. Neuroscience* **1**, 131–136 (1976).
10. Mudge, A. W., Leeman, S. E. & Fischbach, G. D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 526–530 (1979).
11. Yaksh, T. L., Jessell, T. M., Gamse, R., Mudge, A. W. & Leeman, S. E. *Nature* **286**, 155–157 (1980).
12. Jessell, T. M., Mudge, A. W., Leeman, S. E. & Yaksh, T. L. *9th A. Meet. Soc. Neuroscience* Abstr. (1979).
13. Mudge, A. W., Fischbach, G. D. & Leeman, S. E. *7th A. Meet. Soc. Neuroscience* Abstr., 410 (1977).
14. Henry, J. L. *Brain Res.* **114**, 429–451 (1976).
15. Randic, M. & Miletic, V. *Brain Res.* **152**, 196–202 (1978).
16. Phillis, J. W. & Kirkpatrick, J. R. *Can. J. Physiol. Pharmac.* **57**, 887–899 (1979).
17. Mroz, E. A. & Leeman, S. E. in *Methods of Hormone Radioimmunoassay*, 2nd edn, 121–137 (Academic, New York, 1979).
18. Arnold, M. A. & Fernstrom, J. D. *Neuroendocrinology* **31**, 194–199 (1980).
19. Role, L., Leeman, S. E. & Perlman, R. *Neuroscience* (in the press).
20. Greene, L. A. & Shooter, E. M. *A. Rev. Neurosci.* **3**, 353–402 (1980).
21. Barde, Y. A., Edgar, D. & Thoenen, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1199–1203 (1980).
22. Kessler, J. A. & Black, I. B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 649–652 (1980)

© 1981 Nature Publishing Group

23. Otten, U., Goedert, M., Mayer, N. & Lembeck, F. *Nature* **287**, 158–159 (1980).
24. Jessell, T. M., Siegel, R. E. & Fischbach, G. D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5397–5401 (1979).
25. Brockes, J. P., Fields, K. L. & Raff, M. C. *Brain Res.* **165**, 105–118 (1979).
26. Varon, S., Nomura, J. & Shooter, E. M. *Biochemistry* **6**, 2202–2209 (1967).

Inability of tolerant males to sire tolerant progeny

R. N. Smith

Department of Reproductive Biology, Case Western Reserve University, Cleveland, Ohio 44106, USA

Recent data of Gorczynski and Steele suggested that male mice made tolerant of an allogeneic strain by the neonatal injection of hybrid lymphomyeloid cells may, by a presumably genetic mechanism, pass on this operationally similar form of hyporesponsiveness to their progeny[1,2]—the progeny have inherited an acquired characteristic. Their theory, proposed to account for the inheritance of an acquired characteristic, uses a RNA tumour virus that transfers to the spermatogenic germinal epithelium DNA coding for the gene products of the foreign alloantigens. This is possible because the tolerant mouse is chimaeric and contains small numbers of foreign lymphomyeloid cells. Thus, the virus transfers genetic information for transplantation antigens from the chimaeric foreign cells to the germinal epithelium of the host. The progeny of the tolerant male are themselves tolerant because they have received DNA coding for the foreign antigens from paternal sperm. The progeny are tolerant because foreign is now 'self'. I have now repeated these experiments in a strain combination of rats, DA tolerant of PVG, and found no hyporesponsiveness to the PVG strain in the progeny of tolerant DA males.

Neonatally induced transplantation tolerance requires the acceptance of an allogeneic skin graft[3], and such tolerance seems coincidental with the inability of lymphocytes from these skin-graft-tolerant animals to generate cytotoxic lymphocytes against alloantigens of the tolerated strain[4,5]. In addition, the lymphocytes of some tolerant animals are incapable of proliferation in the mixed lymphocyte culture (MLC) when stimulated by alloantigens of the tolerated strain[4,6]. These lymphocytes are also able to transfer adoptively specific tolerance to sub-lethally irradiated syngeneic recipients[4,7]. These four criteria were used to analyse the tolerant state of DA males made neonatally tolerant to PVG, and to analyse the relative responsiveness to PVG alloantigens of the DA progeny of these tolerant males.

DA neonates from three litters (20 pups) were made tolerant of PVG by the injection into the tail vein of 1×10^8 spleen and bone marrow cells from $(PVG\times DA)F_1$ hybrid female donors. Injections of 0.1 ml using a 30-gauge needle were given within 24 h of birth. At 6 weeks of age, all putatively tolerant DA rats—18 of the surviving animals—were grafted with PVG ear skin[8], and of these, 14 accepted PVG grafts for 60 days without evidence of rejection. The skin-graft-tolerant rats were also analysed to determine if their peripheral blood leukocytes (PBL) could be stimulated by PVG alloantigens in the MLC, a sensitive assay used to detect reactivity to allogeneic determinants[9]. DA males that accepted PVG skin grafts for 60 days and had PBL which were not stimulated by PVG alloantigens in the MLC, were allowed to sire one litter with a normal DA female. The tolerant males were cannulated and their thoracic-duct lymphocytes (TDL) analysed in quantitative MLC and cytotoxic lymphocyte culture (CTL) assays for their ability to respond to PVG and third party Lewis alloantigens. In addition, 1×10^8 TDL from these tolerant males were transferred adoptively to sub-lethally irradiated (500 rad) DA recipients to determine if the TDL from tolerant males could transfer unresponsiveness to secondary recipients against PVG antigens[4,7]. Six of the eight males met the four criteria of tolerance to transplantation alloantigens—acceptance of skin grafts, unresponsiveness in the MLC and CTL assays to PVG alloantigens and the ability to transfer suppression (Table 1, Fig. 1). Responses to the Lewis strain, the third party control, were normal, indicating that the tolerance and suppression were specific for PVG antigens.

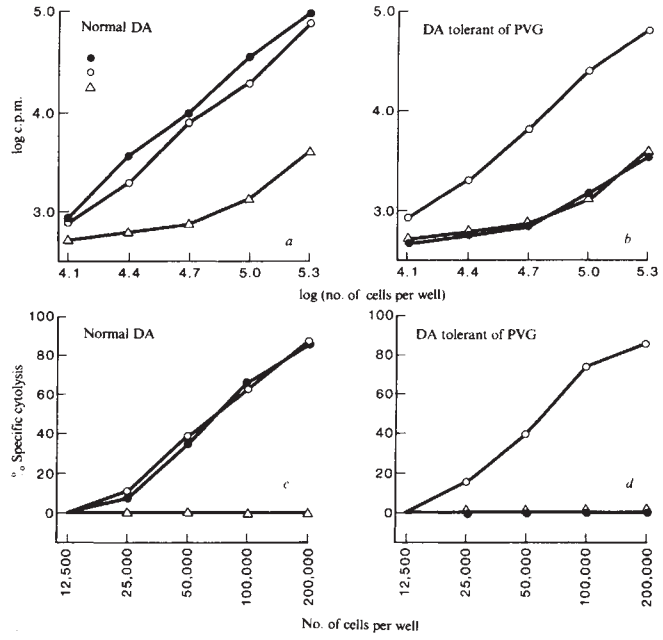


Fig. 1 A representative experiment of the mixed lymphocyte culture (MLC) and cytotoxic lymphocyte culture (CTL) in which responding cells are titrated against a constant number of stimulating cells. The relative estimates of frequency in Table 3 are derived from this type of experiment. Stimulators and targets: △, DA; ●, PVG; ○, Lewis. The MCL was used as a quantitative assay following procedures described elsewhere[10] in which serial dilutions of responder TDL were added to an excess of irradiated stimulator spleen and lymph node cells in a constant volume of 0.2 ml per round-bottomed microtitre well. Cells were cultured in RPMI 1640 containing 5% fetal calf serum (FCS), 2 mM glutamine, 2.5×10^{-5} M 2-mercaptoethanol and antibiotics. The FCS was selected for its ability to support the generation of cytotoxic lymphocytes without high spontaneous proliferation. Stimulator cells were spleen and lymph node cells pooled from one donor and were irradiated with 1,000 rad. The MLC was incubated for 4 days at 37 °C in a humidified CO_2 incubator. During the last 4 h of incubation each well contained 1.0 μCi of ^{3}H-methylthymidine (15 Ci mmol^{-1}), added in 0.01 ml of phosphate buffered saline. The cells were collected on a semi automatic harvester. The same quantitative approach was used to assay the potential cytotoxic activity of a responder population. Cytotoxic activity was measured in wells from titrated MLC responses on day 5. 0.1 ml of culture supernatant was removed from each well and replaced with 0.1 ml of target cells (10^5 per ml). The plates were spun at 500*g* at room temperature for 10 min and incubated at 37 °C for 4–6 h in a humidified CO_2 incubator. To terminate the cytotoxic experiment, plates were placed on ice and 0.1 ml removed for counting. Target cells were 3-day Con A blasts prepared from lymph node cells in complete medium with 2.5 μg ml^{-1} Con A. Target cells ($2–3\times10^7$) were labelled for 1–2 h with 50–100 μCi of ^{51}Cr sodium chromate (Amersham), washed and then resuspended in complete medium.

Half of the progeny (21 DA rats) of the six tolerant males were analysed by quantitative MLC and CTL assays[10], and half were skin grafted with $(PVG\times DA)F_1$ ear skin. Males were grafted with male skin and females with female skin. The data on skin graft rejection (Table 2) indicate that the progeny of the tolerant DA males rejected PVG grafts and third party Lewis grafts at the same rate as did normal DA rats. Thus, there was no evidence of tolerance by this assay.

As the rate at which skin grafts are rejected might not detect subtle and weak hyporesponsiveness, the remaining progeny were analysed to estimate their response to PVG and Lewis alloantigens, as measured by MLC and CTL assays[10]. The number of DA TDL responding to PVG and Lewis alloantigens

0028–0836/81/340767—02$01.00

© 1981 Macmillan Journals Ltd

was titrated to generate a dose–response curve. The advantage of this type of analysis is that displacement along the x-axis represents the relative change in the frequency of cells reactive to the stimulating antigen[10]. The relative frequency between control and experimental groups may be related by comparing the required number of cells, determined by regression, needed to give a certain response. For the MLC assays the number of cells required to give 10,000 c.p.m. ^{3}H-thymidine incorporation was chosen to relate the relative frequency between animals, and the number of cells required for 33% specific cytolysis was chosen as the estimate of frequency for the CTL assay. The data in Table 3 indicate that the relative response of DA TDL from either normal DA rats or DA rats that were the progeny of tolerant males is the same with respect to PVG alloantigens.

These results suggest that, in this strain combination of rats, tolerance or any weak hyporesponsiveness to transplantation antigens cannot be transmitted to the progeny of tolerant rats. Thus, acquired tolerance could not be transmitted genetically. Although these results clearly cannot confirm those of Gorczynski and Steele[1,2], they do not falsify their theory that a virus, which includes in its genome some DNA coding for the tolerated transplantation antigens, could infect spermatogenic germinal epithelium. The resulting sperm would render progeny tolerant because the foreign alloantigen is now self. Presumably, genetic information for other immunological phenomena, for example,

Table 1 Adoptive transfer of tolerance by thoracic-duct lymphocytes from tolerant males

Status of DA male donor	Adoptive transfer assay* (graft survival, days)	
	PVG	Lewis
Rat 1 Tolerant of PVG	14; 48; >100; >100	8; 9; 9; 10
2 Tolerant of PVG	17; >100; >100; >100	9; 9; 9; 10
3 Tolerant of PVG	18; 91; >100; >100	9; 10; 10; 11
4 Tolerant of PVG	21; 53; >100; >100	8; 9; 10; 11
5 Tolerant of PVG	25; 63; >100; >100	9; 9; 10; 10
6 Tolerant of PVG	20; >100; >100; >100	8; 10; 10; 10
Normal DA	9; 10; 10; 11	8; 10; 10; 11
Normal DA	8; 9; 9; 11	9; 9; 9; 10
No cells transferred	17; 19; 24; 24	13; 15; 18; 19
No cells transferred	17; 21; 23; 25	14; 16; 19; 20
No cells transferred	18; 19; 21; 22	15; 18; 19; 19

At the time of cannulation for thoracic-duct lymphocytes (TDL), each DA male tolerant of PVG was 3–4 months old and had not yet rejected a PVG graft 60–90 days after grafting.

* According to Dorsch and Roser[7]. Four DA rats, irradiated with 750 rad (a sublethal dose), were grafted with both PVG and Lewis ear skin and injected intravenously with 5×10^7 TDL from the indicated donor.

anti-idiotypic antibodies, could also be transmitted, and in this way account for the tolerant progeny.

Because viruses show host specificity, that is, preferentially infect certain species or even subgroups of a species, it is possible that mice could transmit genetically acquired tolerance, but that other species, for example, rats, could not.

An additional experiment relevant to the interpretation of the experiments of Gorczynski and Steele is the backcross $(A\times B)F_1$ male × A female. The hybrid male, $(A\times B)F_1$, and the male, A, made tolerant of B are similar in that both have acquired a state of tolerance to B. If such tolerance to B can be transmitted genetically by sperm haploid for A, then less than 50% of the progeny of the backcross should respond to B. As far as I am aware, this is not the case—50% of such progeny are histoin-

Table 2 Skin graft rejection of the progeny of tolerant male rats

Group	No. of animals	Rejection (days) (MST ± 1 s.d. (Range))	
		PVG	Lewis
Normal DA rats	20	7.75 ± 0.72 (7–10)	7.9 ± 0.85 (7–10)
Progeny of DA males tolerant of PVG	21	7.95 ± 0.87 (7–10)	7.81 ± 0.87 (7–10)

All rats were doubly grafted with ear skin from PVG (left flank) and Lewis (right flank).

Table 3 Relative response to PVG alloantigens by thoracic-duct lymphocytes from DA rats

N	Status of DA rats	Slope (mean ± s.d. (range))	No. of cells to give 10,000 c.p.m. (mean ± s.d. (range))
		MLC assay	
10	Normal	1.64 ± 0.15 (1.38–1.84)	49,860 ± 4320 (41,188–57,833)
21	Progeny of males tolerant of PVG	1.70 ± 0.18 (1.41–1.86)	52,495 ± 5950 (40,599–63,495)
		CTL assay	
		Slope (mean ± s.d. (range))	No. of cells to give 33.3% specific lysis
10	Normal	85.6 ± 7.4 (76–94)	49,279 ± 12,226 (45,481–65,234)
21	Progeny of males tolerant of PVG	84.3 ± 6.3 (73–93)	47,321 ± 10,119 (33,258–64,145)

Dose–response curves were done for each animal as described for Fig. 1, and the data analysed by simple linear regression. The relative response to PVG alloantigens was computed for each animal as the number of cells required to incorporate 10,000 c.p.m. of ^{3}H-thymidine or give 33.3% specific lysis of PVG of concanavalin A (Con-A) blasts labelled with ^{51}Cr. Normal and experimental groups are not different statistically by Student's t-test. The correlation coefficient was >0.95 for all regression lines. P >0.2 for the normal and experimental groups in both the MLC and CTL assays.

compatible to B. Because of the generally accepted results of this type of experiment, I believe it is unlikely that 'antigen' is being transferred genetically, and thus this does not explain why tolerant males should pass on their tolerance to their progeny.

Of course, the point of Gorczynski and Steele's work is that genetically transmitted acquired characteristics in mice is not the usual form of inheritance, and that such an event is truly exceptional. It is hoped that the experiments of Gorczynski and Steele[1,2] can be repeated in the same strains of mice which they used, preferably derived from the same stock.

This work was supported by NIH grant HD-13260.

Received 24 April; accepted 2 June 1981.

1. Gorczynski, R. M. & Steele, E. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2871–2875 (1980).
2. Gorczynski, R. M. & Steele, E. J. *Nature* **289**, 678–681 (1981).
3. Billingham, R. E., Brent, L. & Medawar, P. B. *Phil. Trans. R. Soc.* **B239**, 357–414 (1956).
4. Smith, R. N. & Howard, J. C. *J. Immun.* **125**, 2289–2294 (1980).
5. Gorczynski, R. M., Macrae, S. & Till, J. E. *Scand. J. Immun.* **7**, 453–465 (1978).
6. Wilson, D. B., Silvers, W. K. & Nowell, P. C. *J. exp. Med.* **126**, 655–665 (1967).
7. Dorsch, S. & Roser, B. *J. exp. Med.* **145**, 1144–1157 (1977).
8. Smith, R. N. & Powell, A. E. *J. exp. Med.* **146**, 899–904 (1977).
9. Butcher, G. W. & Howard, J. C. *Nature* **266**, 362–364 (1977).
10. Antczak, D. F., Brown, D. & Howard, J. C. *Cell Immun.* **43**, 304–316 (1979).

Two major genes, linked to HLA and Gm, control susceptibility to Graves' disease

Hisamitsu Uno*, Takehiko Sasazuki*, Hajime Tamai† & Hideo Matsumoto‡

* Department of Human Genetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo 113, Japan
† Department of Psychosomatic Medicine, Faculty of Medicine, Kyushu University, Fukuoka 812, Japan
‡ Department of Legal Medicine, Osaka Medical School, Takatsuki 569, Japan

Graves' disease is a multifactorial disease in which immunogenetic as well as environmental factors have important roles. Recently, cumulative evidence has shown that genes controlling immune responses are linked to the MHC (major histocompatibility complex)[1] and/or immunoglobulin allotype genes[2,3]. To identify the genes governing susceptibility to Graves' disease, we have studied 30 Japanese families where more than two first degree relatives were affected with the disease. From genetic analysis of HLA (human MHC) and immunoglobulin allotype of all family members, we present here evidence for the existence of two genes controlling the susceptibility to Graves' disease, one closely linked to HLA-DR, the other linked to the gene coding for the Gm allotype.

0028–0836/81/340768—03$01.00

© 1981 Macmillan Journals Ltd

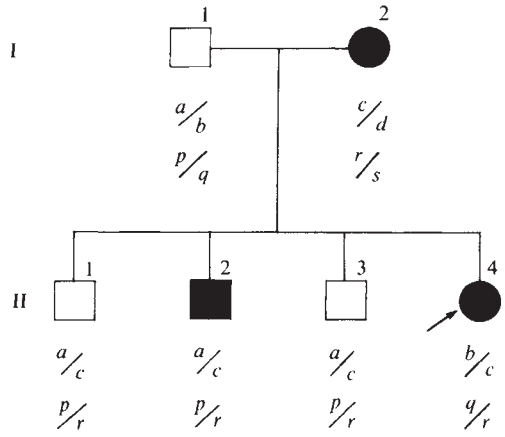


Fig. 1 A family affected with Graves' disease. ● ■, Graves' disease. *a*: HLA-A26, Bw54, Cwl; *b*: HLA-A2, Bw54, –; *c*: HLA-A11, Bw39, –; *d*: HLA-A2, –, –; *p*: Gm(1, 13, 15, 16, 17); *q*: Gm(1, 17, 21); *r*: Gm(1, 3, 5, 13, 23); *s*: Gm(1, 13, 15, 16, 17).

The families comprised 189 members, 71 (37.6%) of whom were affected with Graves' disease (60 female). The diagnosis of Graves' disease was based on the history and signs of hyperthyroidism with diffuse goitre, increased ^{131}I uptake, raised serum thyroid hormone levels, the presence of thyroid autoantibodies, a positive stimulation test to thyrotropin releasing hormone (TRH test) and triiodothyronine suppression test, and histological examination if necessary. All these examinations except the triiodothyronine suppression test were also applied to healthy family members. They were typed for HLA-A, -B, -C and -DR antigens and immunoglobulin allotypes, Gm and Km.

We analysed the segregation of Graves' disease and HLA and Gm haplotypes in 15 selected families. We identified the disease-associated haplotypes of HLA and Gm by determining the haplotypes shared between two affected siblings selected if necessary by random number. We found that all affected siblings except one female shared the disease-associated haplotypes of both HLA and Gm. However, siblings who shared the disease-associated haplotypes did not necessarily suffer from the disease (Fig. 1). Statistical significance of this genetic association between the development of Graves' disease and particular HLA and Gm haplotypes in each family was analysed by the sib-pair method[4] using 62 siblings in the families. Siblings under 10 yr old were excluded from analysis because Graves' disease is very rare below this age (Table 1). A significant association between the development of Graves' disease and particular HLA and Gm haplotypes in each family was observed ($P < 2.0 \times 10^{-3}$). The association was strong and significant in female siblings ($P < 2.0 \times 10^{-3}$), but not significant in male siblings, partly due to the small numbers of male patients in these families. The strong association between the development of Graves' disease and HLA and Gm haplotypes indicated that there might be two genes linked to HLA and Gm, respectively, which had a major influence on the development of Graves' disease. The one affected female sibling who did not share the disease-associated haplotype of HLA with other affected siblings could be explained by a recombination between HLA and an HLA-linked susceptibility gene for Graves' disease. Fourteen healthy siblings who had the disease-associated haplotypes were aged from 12 to 49 yr (mean age 27.6 yr). Because the age of onset of Graves' disease varies widely, they may possibly develop the disease later in life.

Furthermore, in a study of 30 probands in these families, the antigen frequency of HLA-DR5 was significantly increased to 30.0%, compared with 5.0% in control group ($Pc < 3.0 \times 10^{-3}$). The association between Graves' disease and HLA-DR5 indicated that an HLA-linked disease susceptibility gene was in strong linkage disequilibrium with HLA-DR5 (Table 2). There was, on the other hand, no significant association between Graves' disease and any particular immunoglobulin allotype in these probands, indicating the absence of significant linkage disequilibrium between Gm and a Gm-linked disease susceptibility gene.

Table 1 Association between Graves' disease and HLA and Gm haplotypes in 15 families

		Disease-associated haplotypes of HLA and Gm allotype		P
		Positive*	Negative†	
Total sib	Affected	17	1	$<2.0 \times 10^{-3}$
(n = 62)	Unaffected	15	14	
Female sib	Affected	14	1	$<2.0 \times 10^{-3}$
(n = 50)	Unaffected	8	12	
Male sib	Affected	3	0	NS
(n = 12)	Unaffected	7	2	

One affected sibling used to determine disease-associated haplotypes in each family was excluded from this analysis. The P value was calculated by Fisher's exact method. NS, not significant.

* Positive for disease-associated HLA and Gm haplotypes.

† Negative for either or both disease-associated HLA haplotype and Gm haplotype.

The HLA-D region is assumed to be comparable to the I-region of the murine H-2 complex, in which immune response genes (*Ir* genes) and immune suppression genes (*Is* genes) are located[5,6]. Recently, it has become more evident that HLA-linked *Ir* genes or *Is* genes exist in man[7–9]. Analysis of a human population has also revealed a statistical association of Gm allotypes with immune responsiveness to specific antigens[10,11] and with several autoimmune diseases[12,13].

In the murine system, it has been demonstrated that idiotypic determinants are linked to heavy-chain constant region allotypic markers[14,15], and the idiotypic determinants of antigen-specific receptors on lymphocyte membranes are likely to have fundamental roles in immune regulation[16]. The immunoregulatory function of idiotypes was revealed in deliberate immunization experiments[17]. It was therefore assumed that two genes, HLA-linked and Gm-linked, would govern the specific immune response to certain antigen(s) relevant to the development of Graves' disease. The elucidation of the antigen(s), and environmental factors involved in the pathogenesis of Graves' disease will, in turn, help to clarify the immunogenetic factors controlling the susceptibility to Graves' disease.

Table 2 Association between Graves' disease and HLA

HLA-antigens	Graves' disease (n = 30)	Control group (n = 80)	χ^2	Pc	RR
DR5	9(30.0%)	4(5.0%)	13.08	$<3.0 \times 10^{-3}$	8.14
DRw8	11(36.7%)	13(16.3%)	5.82	NS	3.07

DR5 and DRw8 were determined using 11 and 5 sera, respectively, from the 8th International Histocompatibility Workshop[18]. Pc: P multiplied by the number of HLA-DR antigens studied. RR, relative risk.

This is the first description of a strong association between a human disease and not only HLA but also immunoglobulin allotype in a familial analysis.

This study was supported by Grant-in-Aid for Scientific Research 548156 (1980) and for Developmental Scientific Research 487081 (1980) from the Ministry of Education, Japan, and Research Grant (1980) from the Intractable Disease Division, Public Health Bureau, the Ministry of Health and Welfare, Japan.

Received 9 March; accepted 2 June 1981.

1. McDevitt, H. O. & Benacerraf, B. *Adv. Immun.* **11**, 31–37 (1969).
2. Biozzi, B. in *Genetic Control of Immune Responsiveness* (eds McDevitt, H. O. & Landy, M.) 174–178 (Academic, New York, 1972).

© 1981 Nature Publishing Group

3. Smith, S. M., Ness, D. B., Talcott, J. A. & Grumet, F. C. *Immunogenetics* **4**, 221–232 (1977).
4. Penrose, L. S. *Ann. Eugen.* **6**, 133–138 (1935).
5. Benacerraf, B. & McDevitt, H. O. *Science* **175**, 273–279 (1972).
6. Benacerraf, B. & Dorf, M. in *The Role of Products of the Histocompatibility Gene Complex in Immune Responses* (eds Katz, D. H. & Benacerraf, B.) 225–248 (Academic, New York, 1976).
7. Greenberg, L. J., Bradley, P. W., Chopyk, R.-L. & Lalouel, J.-M. *Immunogenetics* **11**, 145–167 (1980).
8. Sasazuki, T., Kohno, Y., Iwamoto, I., Tanimura, M. & Naito, S. *Nature* **272**, 359–361 (1978).
9. Sasazuki, T. *et al. J. exp. Med.* **152**, 297s–313s (1980).
10. Nakao, Y. *et al. New Engl. J. Med.* **304**, 407–409 (1981).
11. Whittingham, S. *et al. Clin. exp. Immun.* **40**, 8–15 (1980).
12. Nakao, Y. *et al. Clin. exp. Immun.* **42**, 20–26 (1980).
13. Farid, N. R., Newton, R. M., Noel, E. P. & Marshall, W. H. *J. Immunogenet.* **4**, 429–432 (1977).
14. Schwartz, M., Lifshitz, R., Givol, D., Mozes, E. & Haimovich, J. *J. Immun.* **121**, 421–426 (1978).
15. Pawlak, L. L., Mushinski, E. B., Nisonoff, A. & Potter, M. *J. exp. Med.* **137**, 22–31 (1973).
16. Jerne, N. K. *Annls Immun. Inst. Pasteur, Paris* **125**, 373–389 (1974).
17. Eichman, K. *Adv. Immun.* **26**, 195–254 (1978).
18. Mickey, M. R. & Terasaki, P. I. in *Histocompatibility Testing 1980* (ed. Terasaki, P. I.) 21–136 (UCLA Tissue Typing Laboratory, California, 1980).

Association between HLA-DR antigens and helper cell activity in the control of dental caries

Thomas Lehner, Jonathan R. Lamb, Kenneth L. Welsh & Richard J. Batchelor

Department of Oral Immunology and Microbiology, Guy's Hospital Medical and Dental Schools, London SE1 9RT, UK
Department of Tissue Immunology, Royal Postgraduate Medical School, The Hammersmith Hospital, London W12 0H5, UK

Dental caries affects the majority of people in developed countries but about 3% of them seem to be resistant. An immunological basis for this resistance has been suggested by the findings that relatively caries-free subjects have increased serum antibody titres[1] and that their T lymphocytes have a greater potential to proliferate on stimulation with streptococcal antigens[2]. Helper cells or their factors[3] are required for formation of antibody to the streptococcal antigen[4]. In this study specific helper factor activity was released by 1–10 ng of the purified streptococcal antigen I/II (SAI/II) from lymphocytes of caries-resistant subjects, compared with 1,000 ng of SAI/II required to release similar helper factor activity from lymphocytes of caries-prone subjects. The low-dose helper activity of lymphocytes was associated predominantly with the HLA-DRw6,1,2,3 cross-reactive groups, whereas the high-dose helper activity was associated with the HLA-DR4-related gene product.

The most important microorganism to be identified as a causative agent of dental caries is *Streptococcus mutans*[5]. A purified protein antigen (SAI/II) isolated from this organism[6,7] induced protection against caries on immunization of rhesus monkeys[8]. SAI/II stimulates monkey peripheral blood lymphocytes to generate T helper cells after *in vitro* incubation for 4 days in a modified Marbrook–Diener system[4]. Further incubation of the helper cells with an optimum dose of SAI/II for 24 h releases antigen-specific helper factor (HF) which stimulates a cooperative antibody response by normal mouse spleen cells to dinitrophenylated SAI/II (DNP–SAI/II). We have now modified this entirely *in vitro* method of generating both helper cells and factors, by assuming that *in vivo* natural immunization in man and artificial immunization in rhesus monkeys should lead to generation of helper cells. We have then used 1–10,000 ng of SAI/II to stimulate the relase of HF from the putative helper cells.

We selected a population of 24 young subjects (mean age 23 yr), 12 of whom showed a low DMFS index (decayed, missing, filled tooth surfaces) of 0–3 (mean ± s.e., 1.0 ± 0.4), and the remaining 12 showing a high DMFS index of 5–35 (18.8 ± 3.0). Lymphocytes from the caries-resistant subjects released HF on addition of 1–10 ng SAI/II, whereas lymphocytes from

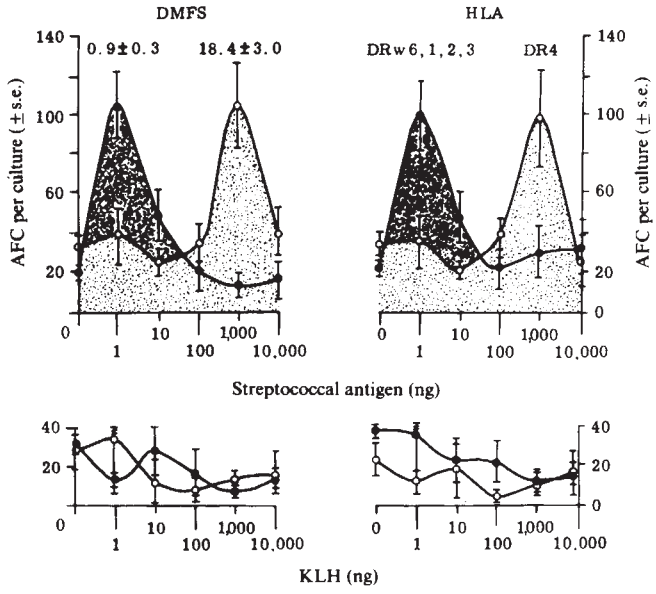


Fig. 1 The effect of dose of streptococcal antigen I/II (SAI/II) on the T-cell helper function in caries-resistant and caries-prone subjects. Human peripheral blood lymphocytes were separated by Ficoll–Isopaque density gradient[18]. 5×10^6 viable cells per ml were cultured with 1, 10, 100, 1,000 and 10,000 ng of SAI/II[6,7] or keyhole limpet haemocyanin (KLH) for 24 h in Marbrook flasks[4]. The cell-free supernatant (HF) was collected, Millipore filtered and 1 μl of the HF was assayed in Marbrook–Diener culture flasks with 10^7 unprimed B10.BR spleen cells and 0.1 μg ml^{-1} DNP-haptenated SAI/II or TNP–KLH[4]. The anti-DNP antibody-forming cells (AFC) were assayed on day 4 in triplicate using the modified Cunningham assay[19]. The results are expressed as the number of AFC per culture ± s.e.m. and are given for the low DMFS (decayed, missing and filled surfaces; $n = 12$) and high DMFS group ($n = 12$) on the left, and DRw6,1,2,3 ($n = 14$) and DR4 group ($n = 10$) on the right. Lymphocytes responding to 1–10 ng SAI/II (DMFS 1.0 ± 0.4; HLA-DRw6,1,2,3) are shaded and those responding to 1,000 ng SAI/II (DMFS 18.8 ± 3.0; HLA-DR4) are stippled.

caries-prone subjects required 1,000 ng SAI/II for a similar release of HF, assessed by the antibody-forming cell assay (Fig. 1). Indeed, 11/12 (92%) of the caries-resistant, compared with 3/12 (25%) caries-prone subjects, released a significant amount of HF from lymphocytes stimulated with 1–10 ng SAI/II ($P = 0.0027$ by Fisher exact test). In contrast, 8/12 (68%) of the caries-prone subjects and 0/12 caries-resistant subjects released a significant amount of HF from lymphocytes stimulated with 1,000 ng SAI/II ($P = 0.0013$). Significant HF was taken as the mean + 3 s.d. (79) antibody-forming cells resulting from the HF activity released by lymphocytes in the absence of SAI/II in 24 subjects.

The striking difference in the dose response to the SA between the disease-resistant and prone subjects was then analysed in relation to the HLA-A,B,C and DR(w) antigens[9]. There were no obvious differences with the HLA-A,B and C antigens but a significant relationship was observed with the HLA-DR antigens (Fig. 1). Stimulation of lymphocytes with 1–10 ng SAI/II released significant amounts of HF activity in 7/9 subjects with DRw6, as compared with 1/10 with DR4 ($P = 0.0097$). Furthermore, a smaller group of five DR1,2,3-typed lymphocytes also showed a significantly greater number of lymphocytes responding to the lower dose of SAI/II than the DR4-typed lymphocytes ($P = 0.0039$). It seems, therefore, that the cross-reacting DRw6,1,2,3 antigens[10] are associated wih helper cells responding to low doses of SAI/II (12/14) whereas the DR4 antigens are not (1/10; $P = 0.00074$). However, lymphocytes with DR4 responded to 1,000 ng SAI/II in 6/10 (60%) subjects, compared with 3/14 (14%) subjects with lymphocytes bearing DRw6,1,2,3 antigens ($P = 0.0566$). Note that DR4 seems to be functionally dominant, because in all 10

0028–0836/81/340770—03$01.00
© 1981 Macmillan Journals Ltd

Table 1 Reproducibility of the dose responses of antibody-forming cells of lymphocytes

Subject	DRw	No. of times tested	ng of streptococcal antigen I/II 0	1	10	100	1,000	10,000
1	6	4	21.8 (2.8)	127.2 (7.2)	15.8 (5.6)	31.3 (14.4)	16.8 (7.1)	17.5 (5.9)
2	6	3	15.3 (2.9)	84.7 (30.4)	93 (15.1)	14.3 (9.8)	15.7 (8.7)	12.3 (4.7)
3	4	5	22 (3.3)	34 (5.6)	40 (20)	34.3 (1.3)	136.6 (12.6)	47.7 (3.9)
4	4	3	15.3 (2.9)	15 (4)	45 (15)	43 (12)	133.3 (20.3)	60 (21.6)

Results are the mean (±s.e.) of antibody-forming cells (AFC). The AFC of human peripheral blood lymphocytes were determined to five concentrations of streptococcal antigen I/II, as described in Fig. 1 legend. This was performed on three, four or five occasions on lymphocytes from the same four donors.

subjects in whom DR4 was associated with DRw6,1,2,3 the lymphocytes responded to the high dose of SAI/II. The reproducibility of the effect of increasing doses of SAI/II on lymphocytes was tested on three to five different occasions in four subjects (Table 1). The maximum number of antibody-forming cells resulted from 1–10 ng SA for DRw6 or 1,000 ng for DR4 on each of the 15 occasions tested; the variations between different tests in the same subjects were small, as shown by the standard errors of the means.

A parallel study of HF activity of *in vivo* immunized, caries-resistant monkeys with SAI/II revealed that lymphocytes from these monkeys responded to 1–10 ng of SAI/II but not to keyhole limpet haemocyanin (KLH; Fig. 2). However, caries-prone monkeys which were not immunized showed no response to any of the antigen concentrations. The results of naturally acquired immunity in man and artificially acquired immunity in monkeys suggests that protection against dental caries in both primates is associated with T helper cells capable of responding to 1–10 ng SAI/II; this might be controlled in man by the cross-reactive DRw6,1,2,3 gene products. However, the caries-prone subjects have T cells which require a 1,000 times higher dose of SAI/II to respond with a HF activity comparable to cells of the caries-resistant subjects, and DR4 may be one of the gene products controlling this response. As such a high dose is unlikely to be reached in biological conditions, the end result might be similar to that found in monkeys which failed to respond to any of the antigen doses.

The specificity of the HF to DNP–SAI/II and characterization of HF have been described for rhesus monkeys elsewhere[4,11]. Corresponding specificity studies have been carried out on human lymphocytes. Culturing human or monkey lymphocytes with 1–10,000 ng KLH and using trinitrophenyl (TNP)–KLH in the cooperative culture had no effect on the number of antibody-forming cells (Figs 1, 2). To exclude further the possibility that the SAI/II might stimulate a nonspecific helper factor, the factor was added to the cooperative culture with TNP–KLH. This had no effect on the number of antibody-forming cells (22.5±7.4, $n=4$), as compared with the same four factors added to the cooperative culture with DNP–SAI/II, which increased the antibody-forming cells to 115.0 (±2.2).

Immunoadsorption of HF on Sepharose 4 beads coated with SAI/II bound the HF activity but did not do so when the beads were coated with KLH (in preparation). Furthermore, as with monkey HF[11], human HF was bound by antisera to anti-HF, but not anti-suppressor factor antisera. The factor also had DR region-encoded determinants, as insolubilized monoclonal anti-DR framework antibody (DA2) bound HF. Moreover, a monoclonal antiserum to DRw6,1,2 (Genox 353) absorbed out all the helper activity of DRw6 lymphocytes, without any effect on the helper activity from DR4 lymphocytes (in preparation). The factor binds IgM but not IgG or IgA antisera, as has been found in mouse and monkey helper factors[11–13]. The significance of IgM has not been adequately explained, but it has been interpreted as a cross-reaction with some region of the μ chain. It seems, therefore, that HF prepared from human lymphocytes has, in addition to an antigen-binding site, a 'constant' region and determinants encoded by the HLA-D region, as has been found in HF prepared from mouse lymphocytes[14].

The finding that T helper cells respond to either low or high doses of antigen in man has no direct parallel in animal experiments. T-cell involvement has been observed in low responder mice, for in addition to the decreased antibody formation there is a decrease in delayed hypersensitivity and *in vitro* T-cell proliferative response[15]. Furthermore, immunization with an antigen under *Ir* gene control showed that in high responder mice the number of antigen-specific T cells increases, whereas no change occurs in the low responders[16]. The association of low-dose T helper cell responses in man with the DRw6,1,2,3 cross-reactive antigens and high-dose helper responses with DR4-related antigen might be an expression of gene control. However, *in vitro* secondary responses to KLH revealed that human T-cell monocyte interactions are governed by HLA-D region antigens[17]. It is therefore possible that the functional high- and low-dose responses to SAI/II which are associated

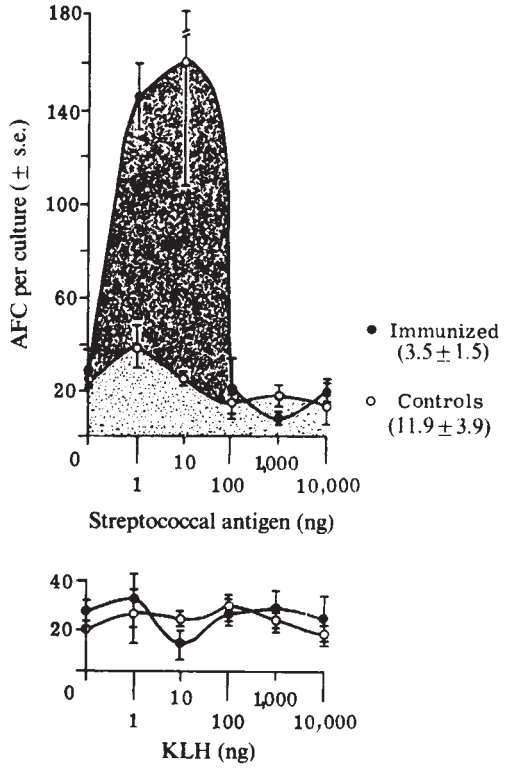


Fig. 2 The effect of streptococcal antigen I/II (SAI/II) T cell helper function in immunized and control monkeys. Rhesus monkey peripheral blood lymphocytes were separated and cultured with 1, 10, 100, 1,000 and 10,000 ng of SAI/II or KLH. HF was prepared and the AFC were assayed and expressed as described for human lymphocytes in Fig. 1 legend. The results are given for six monkeys immunized with SAI/II and with a caries score of 3.5 (±1.5) and six nonimmunized monkeys with a score of 11.8 (±3.9). Lymphocytes responding to 1–10 ng of SAI/II (DMFS of 3.5) are shaded and there are no lymphocytes responding to 1,000 ng of SAI/II among the control monkeys.

© 1981 Nature Publishing Group

with the HLA-D region may be related to monocyte antigen-binding cells and not necessarily to T cells. This possibility will be examined, although preliminary results with radiolabelled SAI/II suggest that selective binding of small doses of SA to HLA-DRw6,1,2,3 cells and high doses to HLA-DR4 cells are related to T cells. Whichever set of cells is involved, it seems that the aims of immunization should be to induce with very small doses of protein antigen T helper cells which cooperate with B cells in the formation of high-avidity antibodies, required in protection against the microorganisms.

We thank Dr M. Crumpton for antisera and Miss Janet Avery for technical assistance. This investigation was carried out under a MRC project grant.

Received 1 April; accepted 16 June 1981.

1. Challacombe, S. J. & Lehner, T. *J. dent. Res.* **55**, C139–C148 (1976).
2. Lehner, T., Challacombe, S. J., Wilton, J. M. & Ivanyi, L. *Archs oral Biol.* **21**, 749–753 (1976).
3. Feldman, M. *J. exp. Med.* **136**, 737 (1972).
4. Lamb, J. R., Kontiainen, S. & Lehner, T. *J. Immun.* **124**, 2384–2389 (1980).
5. Fitzgerald, R. J. & Keyes, P. H. *J. Am. med. Ass.* **61**, 23–33 (1960).
6. Russell, M. W. & Lehner, T. *Archs oral Biol.* **23**, 7–15 (1978).
7. Russell, M. W., Bergmeier, L., Zanders, E. D. & Lehner, T. *Infect. Immunity* **28**, 486–493 (1980).
8. Lehner, T., Russell, M. W. & Caldwell, J. *Lancet* **i**, 995–996 (1980).
9. Welsh, K. I. & Batchelor, J. R. in *Handbook of Experimental Immunology* 3rd edn (ed. Weir, D. M.) 35.1–35.20 (Blackwell, Oxford, 1978).
10. Bodmer, J. G. *Br. med. Bull.* **34**, 233–240 (1978).
11. Zanders, E. D., Lamb, J. R., Kontiainen, S. & Lehner, T. *Immunology* **41**, 587–596 (1980).
12. Feldman, M. & Basten, A. *Nature* **237**, 13 (1972).
13. Taniguchi & Tada *J. Immun.* **113**, 1757 (1974).
14. Kontiainen, S. & Feldman, M. *Thymus* **1**, 59–64 (1979).
15. Benacceraf, B. & McDevitt, H. O. *Science* **175**, 272–277 (1972).
16. Hämmerling, G. I. & McDevitt, H. O. *Transplantn Proc.* **5**, 179–182 (1973).
17. Rodey, G. E., Luehrman, L. K. & Thomas, D. W. *J. Immun.* **123**, 2250–2254 (1979).
18. Wilson, B. J. & Kocvara, H. *J. Immun. Meth.* **9**, 67–68 (1975).
19. Cunningham, A. & Szenberg, A. *Immunology* **14**, 599–600 (1968).

Cross-reactivity between Thy-1 and a component of intermediate filaments demonstrated using a monoclonal antibody

R. Dulbecco*, M. Unger, M. Bologna†, H. Battifora†, P. Syka & S. Okada

The Salk Institute, 10010 N. Torrey Pines Rd, La Jolla, California 92037, USA and *Department of Pathology and Medicine, University of California at San Diego, La Jolla, California 92037, USA

Monoclonal antibodies[1] are powerful tools for identifying antigenic molecules because they recognize a single antigenic site. In contrast, conventional immune sera contain mixtures of antibody molecules with a large variety of specificities. Some of these specificities may be cross-reactive with antigenic sites present on unrelated molecules, thus tending to decrease the overall recognition specificity. Such a cross-reactivity should be observed with some monoclonal antibodies, in which it might be expected to be both strong and specific (that is, directed against a restricted number of targets[2]). We report here such a cross-reactivity in a monoclonal antibody that recognizes equally well both Thy-1 (ref. 3) and a component of intermediate filaments, probably vimentin[4,5].

The hybridoma producing the monoclonal antibody in question was isolated by Lake and Clark[6]. This antibody is an anti-Thy-1.1 mouse IgM designated T11A9e, which binds to Thy-1 at the surfaces of various cells, for example, the fusiform derivatives of the mammary cells RAMA[7,8] and RANI[9] or the F2408 rat fibroblasts[10] (Fig. 1*a*). In the present experiments, the antibody used was either in the form of serum obtained from mice in which the T11A9e hybridoma had given rise to a tumour after subcutaneous injection, ascitic fluid after intraperitoneal cell injection, or supernatant of the hybridoma cultures. The IgM in the supernatant was partially purified and concentrated about 30 times by retention on an Amicon XM300 filter, followed by sizing in a Biogel A5M column. All these preparations had identical specificities.

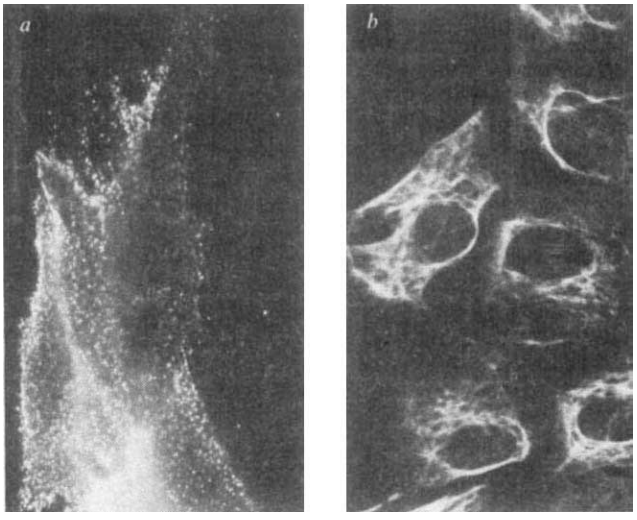


Fig. 1 Immunofluorescence of T11A9e serum with F2408 cells. *a*, A glass cover slip on which cells had grown was washed in phosphate-buffered saline (PBS) containing 0.01% Na azide and was then covered with antiserum diluted 1/50 in the PBS–azide saline. After 30 min, it was washed three times with PBS–azide, then overlayered with a 1/50 solution of fluorescein-labelled goat anti-mouse immunoglobulin (Antibodies, Inc.) for 30 min, washed again three times and mounted. Similar staining, of slightly reduced intensity, was obtained if the cells were fixed in 3% formaldehyde for 20 min. *b*, The cover slip was fixed in 3% formaldehyde for 20 min, rinsed in PBS–azide containing 0.01% thymerosal, dipped in cold methanol and then acetone (30 s each), hydrated in PBS–azide and then stained as in *a*, except that the three washings after each antibody application were for 10 min, 2 h and 10 min. Similar results were obtained if formaldehyde fixation was omitted. The disappearance of the Thy-1 surface dots was observed for all antisera. Photographed using a Zeiss microscope with epi-illumination. *a*, ×1,300; *b*, ×900.

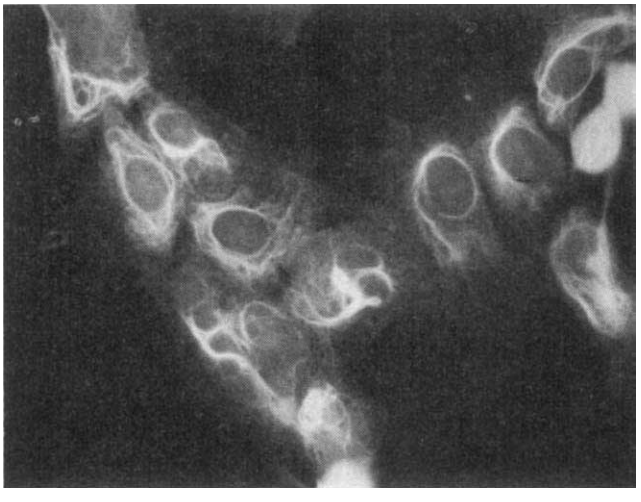


Fig. 2 Immunofluorescence of F2408 cells treated with colchicine. Cover-slip cultures were grown for 18 h in the presence of colchicine, 20 $\mu g\ ml^{-1}$, then stained with T11A9e serum as described for Fig. 1*b*. ×700

† Permanent addresses: Department of Pathology, University of L'Aquila Medical School, L'Aquila, Italy (M.B.) and Department of Surgical Pathology, Northwestern University Medical School, Chicago, Illinois 60611, USA (H.B.).

0028-0836/81/340772—03$01.00

© 1981 Macmillan Journals Ltd

Immunofluorescence experiments showed that T11A9e stains, in addition to surface Thy-1 in unfixed cells, a system of filaments in methanol–acetone-treated cells. Stained filaments were seen in the fusiform but not in the epithelial derivatives of mammary cells, and in the F2408 fibroblasts (Fig. 1*b*). They were also stained in cells that do not contain a Thy-1.1 antigen on their surfaces, such as BALB/c macrophages or BALB/c 3T3 fibroblasts, which have Thy-1.2. In all cells the pattern resembled that of intermediate filaments stained by antibodies to vimentin. In fact, in double immunofluorescence experiments, an identical network was stained in F2408 cells using either T11A9e or a rabbit antiserum to vimentin (Fig. 2). In cells pretreated with colchicine (20 $\mu g\ ml^{-1}$) for 12 h, the filaments stained by T11A9e showed the behaviour reported for intermediate filaments[4] (Fig. 3). To exclude the possibility that the ability to stain these filaments was due to an accidental contamination of the hybridoma culture with hybridoma cells producing a different antibody, the T11A9e hybridoma was recloned. This recloned hybridoma produced an antibody with the same properties.

To determine whether the same antibody molecules stained Thy-1 and intermediate filaments, a T11A9e antiserum was absorbed with cells of two Thy-1.1-positive thymoma cell culture lines, or their Thy-1 negative derivatives[11]. The ability to stain the filaments was absorbed by the Thy-1-positive cells but not by their negative variants; it was not absorbed by Thy-1.2 cells, whether positive or negative for Thy-1 (Fig. 4).

To test whether T11A9e antibody recognizes intermediate filaments through its Fc segment, the T11A9e antibody was preincubated for 30 min at room temperature with various concentrations of Thy-1.1 purified from rat brain (provided by I. Trowbridge). The mixture was then used to stain either Thy-1 at the surface of unfixed F2408 cells, or intermediate filaments in acetone-treated F2408 cells. Thy-1 at a concentration of 30 $\mu g\ ml^{-1}$ completely inhibited both types of staining. The Thy-1 concentrations at which the proportion of surface-stained cells or the intensity of filament staining (as judged from visual observation) was reduced to 50% were similar, $\sim 1\ \mu g\ ml^{-1}$. Therefore both target molecules are recognized by the same antibody site.

In addition to T11A9e, two out of 13 monoclonal anti-Thy-1 antibodies also stain intermediate filaments; these are T12G11c (IgM) and T32B11e (IgG) (both from P. Lake and E. Clark). The others give only a faint uniform fluorescence, probably due to surface Thy-1. Of the positive antibodies, T12G11c may be from the same lymphocyte clone as T11A9e, but T32B11e is likely to be different. Two conventional anti-Thy-1 rabbit antisera, obtained using purified Thy-1 from rat brain or from mouse thymocytes, did not stain the filaments.

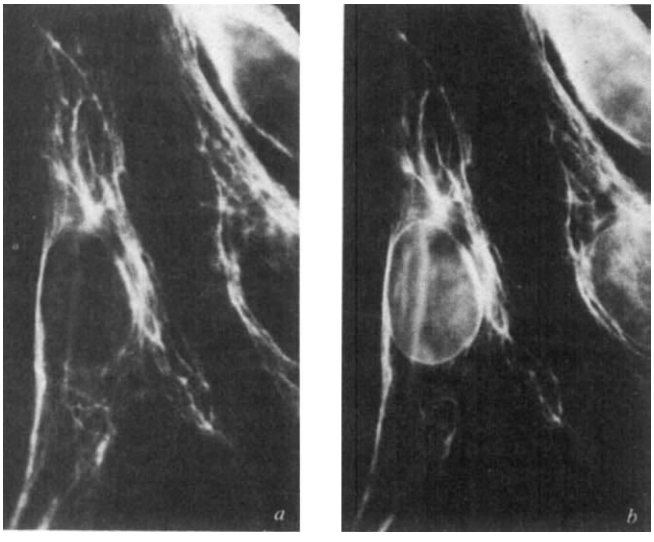


Fig. 3 Double immunofluorescence of F2408 cells with T11A9e and anti-vimentin serum. Immunofluorescence was carried out as described for Fig. 1*b*. *a*, The first antibody was a rabbit anti-vimentin serum and the second antibody was rhodamine-conjugated goat anti-rabbit immunoglobulin. *b*, The first antibody was T11A9e serum and the second, fluorescein-conjugated goat anti-mouse immunoglobulin. Photographs were taken using the appropriate filters for detection of fluorescence by rhodamine (*a*) or fluorescein (*b*). ×1,400.

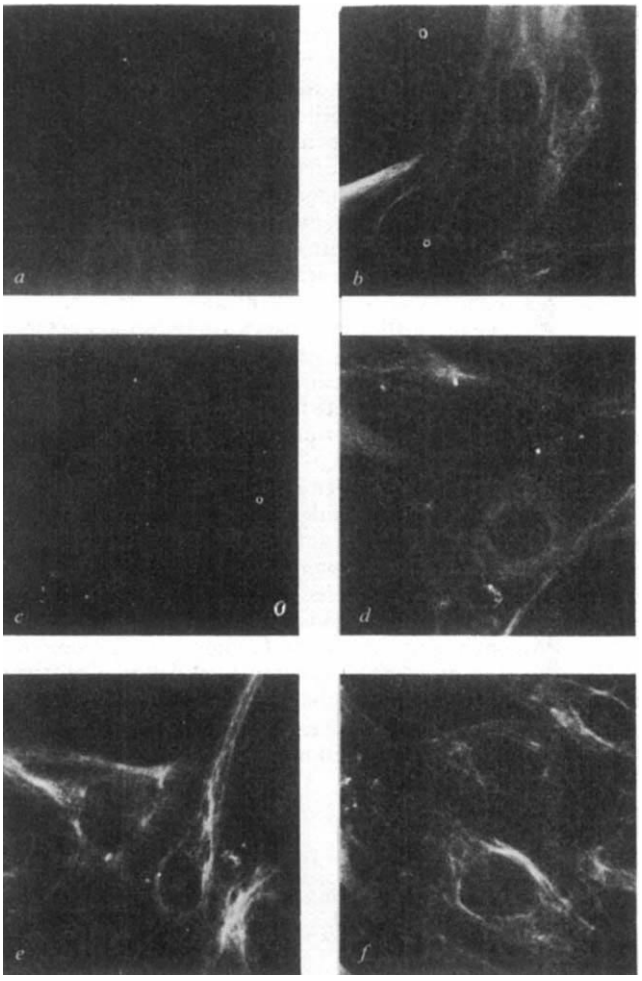


Fig. 4 Absorption of T11A9e serum with Thy-1-positive cells and their negative mutants. 100 μl of serum diluted 1/100 in PBS were absorbed in each case with 5×10^8–5×10^9 cells of various mouse thymoma lines[12] (provided by R. Hyman). After 1 h at room temperature with frequent shaking the mixtures were centrifuged at 2,000 r.p.m. for 10 min. The supernatants were used in immunofluorescence experiments with F2408 cells as described for Fig. 1*b*. Absorbing cells: *a*, BW5147.3 · Thy-1.1$^+$; *b*, BW 5147.3 · Thy-1.1$^-$e; *c*, BW5147 · Thy-1.1$^+$; *d*, BW 5147 · Thy-1.1$^-$a; *e*, S49 · Thy 1.2$^+$; *f*, S49 · Thy 1.2$^-$a. The fields were photographed and printed using identical exposures. ×800.

To identify the protein to which the T11A9e antibody binds in the intermediate filaments, cytoskeleton proteins of F2408 cells were isolated as previously described[12], labelled with ^{125}I then immunoprecipitated with T11A9e antiserum, using a second antibody. Separation by polyacrylamide gel electrophoresis and autoradiography did not reveal any band of precipitated material. This antibody also fails to precipitate surface Thy-1. These failures may reflect a low affinity of the antibody. In spite of the lack of immunoprecipitation it is likely that T11A9e antibodies bind to vimentin because the T11A9e immunofluorescence of intermediate filaments is suppressed by anti-vimentin antiserum.

Because serum from mice inoculated with another isolate of the same hybridoma (T11A9a, no longer available) had a strong *in vitro* anti-doming effect on mammary cells[13], we examined the possible relationship of this effect with our results. We found

© 1981 Nature Publishing Group

that the anti-doming effect was not removed by absorption to Thy-1.1-positive cells and was therefore unrelated.

We do not know whether the cross-reactivity we have observed is due to the proteins or possibly to some common carbohydrate. If the former is correct, it might reveal either a relatedness of the two proteins or a chance similarity of antigenic sites. In the latter case, our observation would be an example of the cross-reactivity of monoclonal antibodies predictable on theoretical grounds[2]. The frequency of cross-reacting antibodies is unknown. When mice and rats were immunized with Thy-1, antibodies showing the cross-reaction we described occurred in 3 out of 13 monoclonal antibodies examined, but may not represent the true proportion because two of the antibodies may have a common origin. No general statement about the frequency of cross-reactivities can be made on the basis of these results or of past experience with immune sera. This remains to be established in further experiments using monoclonal antibodies.

We thank Drs I. S. Trowbridge and R. Hyman for advice, Drs P. Lake, E. Clark, I. S. Trowbridge, A. Williams and E. Weissman for supplying monoclonal antibodies, and J. Singer for the antivimentin antiserum. This work was supported by USPHS grant 1-RO1 CA21993, awarded by the NCI, DHHS, and by grants from the Hammer Foundation, the Educational Foundation of America, and the Pardee Foundation. The research was conducted in part by the Clayton Foundation, California Division. R.D. is a Senior Clayton Foundation Investigator and M.U. a Clayton Foundation Investigator. M.B. was a recipient of the Angela Bossolasco Fellowship of the Accademia de Lincei (Italy).

Received 13 April; accepted 24 June 1981.

1. Williams, A. F., Galfre, G. & Milstein, C. *Cell* **12**, 663–673 (1977).
2. Milstein, C. & Lennox, E. in *Curr. Topics dev. Biol.* **14**, 1–32 (1980).
3. Williams, A. F. in *Contemporary Topics in Molecular Immunology* (eds Porter, R. R. & Ada, G. L.) 83–114 (Plenum, New York, 1977).
4. Franke, W. W., Schmied, E., Osborn, M. & Weber, K. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5034–5039 (1978).
5. Hynes; O. & Destree, A. T. *Cell* **13**, 151–163 (1978).
6. Lake, P. & Clark, E. A. *Eur. J. Immun.* **9**, 875–886 (1979).
7. Bennett, D. C., Peachey, L. A., Durbin, H. & Rudland, P. S. *Cell* **15**, 283–298 (1978).
8. Rudland, P. S., Bennett, D. C., Ritter, M. A., Newman, R. A. & Warburton, M. J. in *Control Mechanisms in Animal Cells* (eds Jemenez de Asua, L. *et al.*) 341–365 (Raven, New York, 1980).
9. Dulbecco, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 2345–2349 (1981).
10. Freeman, A. E., Igel, H. J. & Price, P. J. *In Vitro* **11**, 107–116 (1975).
11. Trowbridge, I. S., Hyman, R. & Mazauskas, C. *Cell* **14**, 21–32 (1978).
12. Franke, W. W., Schmid, E., Osborn, M. & Weber, K. *J. Cell Biol.* **81**, 570–580 (1979).
13. Dulbecco, R., Bologna, M. & Unger, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1848–1852 (1979).

Enzymatic synthesis of deshexapeptide insulin

Q. P. Cao, D. F. Cui & Y. S. Zhang

Shanghai Institute of Biochemistry, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai, China

The semisynthesis of deshexapeptide insulin (DHI) from desoctapeptide insulin (DOI) reported earlier[1] showed that DHI was biologically active. However, the product was not quite homogeneous, mainly due to the esterification side reaction and the following saponification[2]. Recently, the catalytic action of trypsin was successfully used in the preparation of human insulin from porcine insulin[3]. It was also used in the preparation of insulin analogues with substitutions of B24 or B25 phenylalanine by leucine[4,5]. Here, we report the synthesis of DHI from DOI and glycylphenylalanine through the catalytic action of trypsin. The product, purified by gel filtration and DEAE-Sephadex ion-exchange chromatography, showed insulin activity of 7.5 IU mg^{-1} *in vivo* and could be obtained in crystalline form.

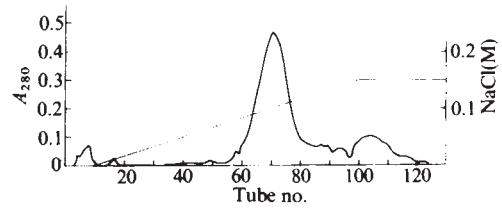


Fig. 1 Isolation of DHIOBu^t from DOI by ion-exchange chromatography on a DEAE-Sephadex A-25 column (1 × 25 cm). 50 mg of sample were applied to the column, eluted using 50 ml of the initial 0.05 M Tris buffer, *p*H 7.6, with 40% isopropanol and then eluted through a salt gradient (from 200 ml of initial buffer containing 0.15 M NaCl to 200 ml of initial buffer alone). The eluent was collected in ~4-ml fractions. Active fractions were pooled, desalted by dialysis and lyophilized. The main peak is DHIOBu^t and the last peak, DOI.

The conditions for the enzymatic synthesis of DHI were different in certain aspects from those reported by Inouye *et al.*[3]. The alkali-labile methylsulphonylethoxycarbonyl group (Msc) was used[6] instead of the *t*-butyloxycarbonyl group (Boc). The *t*-butyl ester (OBu^t) was not affected by the removal of Msc[6], and DHIOBu^t had one less negative charge than DOI; thus, they could be easily separated from each other by ion-exchange chromatography.

120 mg of (Msc)$_2$DOI and Gly-Phe-OBu^t in excess (molar ratio = 1:60) were suspended in 0.16 ml of dimethyl sulphoxide (DMSO) and 1.08 ml of dimethylformamide (DMF) were added to ensure complete solubility. The solution was warmed to 37 °C and 0.96 ml 0.5 M Tris buffer, *p*H 8, added dropwise and the *p*H maintained at 7.8–8.0 by continual addition of triethylamine. Crystalline ovine trypsin was added and the solution incubated at 37 °C for 20 h. The total amount of enzyme used (12 mg) was added in three portions: at the start and after 2 and 6 h of incubation. At the end of the incubation, 40 ml of water were added and the *p*H adjusted to 4.2 using 1 M HCl. After cooling, the precipitate was collected by centrifugation, washed in water (*p*H 4.2) and lyophilized; 95.5 mg of dry powder were obtained. This powder was suspended in 4.78 ml of a solvent mixture, dioxane–methanol–water (3:0.5:3.5) and dissolved by adding a small amount of 2 M NaOH. The solution was cooled to 0 °C and 2 M NaOH at 0 °C added until the total amount was 0.68 ml. The deblocking reaction was carried out at 0 °C for 2 min and immediately stopped by the addition of acetic acid. The solution was then applied to a Sephadex G-50 column and eluted with 1 M acetic acid to remove high-molecular weight polymers. DHIOBu^t and DOI were separated by DEAE-Sephadex chromatography (see Fig. 1). The purified DHIOBu^t

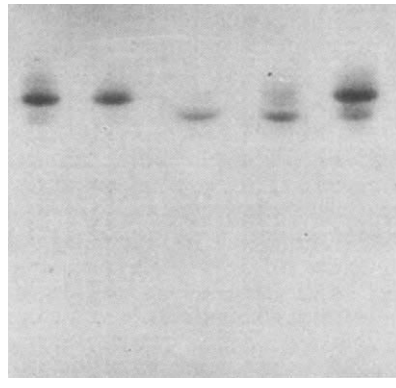


Fig. 2 Polyacrylamide gel electrophoresis of, from left to right: crude DHIOBu^t, purified DHIOBu^t, purified DHI, DOI and crude DHIOBu^t; the lower end is the anode.

0028–0836/81/340774—02$01.00

© 1981 Macmillan Journals Ltd

was treated with trifluoroacetic acid (TFA, 2.5 mg ml^{-1}) at 0 °C for 1 h. TFA was then removed by evacuation and the residue washed twice with ethyl acetate. 10 mg of the product were further purified by DEAE-Sephadex A-25 chromatography to remove any unreacted DHIOBu^t and 7 mg of purified DHI were obtained. The purified DHI was dissolved in 0.01 M HCl (1.5 mg ml^{-1}) and dialysed against citrate buffer containing 25% acetone (2% citric acid, 50 ml; 6.5% zinc acetate, 1 ml; acetone, 25 ml; adjusted to *p*H 5.7–5.8 with NaOH and diluted to 100 ml). DHI crystals appeared after standing overnight.

The biological activity of purified DHI was determined in a mouse convulsion assay (according to British Pharmacopoeia) as 7.5 IU mg^{-1}. DHIOBu^t also possessed insulin activity, thus it was easily detected and collected by DEAE-Sephadex chromatography. However, its activity is about four times lower than that of DHI. The activity of DOI is less than 1% compared with insulin.

After DEAE-Sephadex chromatography (Fig. 1), the DHIOBu^t peak yielded 29.8 mg and the DOI peak, 4.9 mg; a ratio of 86 : 14. Polyacrylamide gel electrophoresis[7] showed that

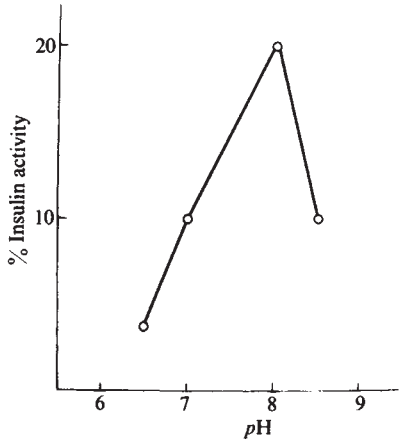


Fig. 3 The optimum *p*H for the enzymatic synthesis of DHI from $(Boc)_2$ DOI and Gly-Phe-OBu^t.

most of the DOI was converted into DHIOBu^t (Fig. 2). DOI moved faster towards the anode than DHIOBu^t. When OBu^t was removed, DHI occupied the same position as that of DOI. Purified DHI gave one major band after electrophoresis. Its amino acid composition was determined as: His, 2.4 (2); Arg, 0.9 (1); Asx, 3.0 (3); Glx, 6.3 (7); Thr, 1.1 (1); Ser, 2.7 (3); Gly, 3.8 (4); Ala, 1.1 (1); Val, 3.4 (4); Ile, 1.7 (2); Leu, 5.1 (6); Tyr, 2.6 (3); Phe, 1.9 (2) $(Cys)_2$ not determined.

In enzymatic synthesis, co-solvent must be used—we used Tris buffer containing 50–60% DMF with a small amount of DMSO to increase the solubility. In tryptic hydrolysis, the optimal *p*H was found to be around 8 (Fig. 3). Inouye *et al.*[3] used a *p*H of 6.5 for their synthesis. We have compared the enzymatic synthesis of DHI from $(Boc)_2$DOI and Gly-Phe-OBu^t in solutions of different *p*H, and found that the optimum *p*H was 8. The yield of the enzymatic reaction of $(Msc)_2$DOI and Gly-Phe-OBu was 74%, calculated from the amount of DHIOBu^t and DOI obtained by DEAE-Sephadex chromatography and the small portion of DOI polymers in gel filtration. This result could be attributed to several factors: (1) An increase in the equilibrium constant of synthesis due to the presence of co-solvent, which decreases the equilibrium constant of proton transfer from the carboxyl to the amino group of the reactants, as reported by Homandberg *et al.*[8]. (2) The high concentration of reactants and the large excess of Gly-Phe-OBu^t drive the reaction towards synthesis according to the law of mass action. (3) The *p*H (8) was optimal for the catalytic action of trypsin. (4) Both reactants are very soluble at this *p*H.

We thank Dr D. Saunders for supplying Msc *N*-hydroxysuccinimide.

Received 5 January; accepted 18 June 1981.

1. Shanghai Insulin Research Group. *Scientia sin.* **16**, 61–70 (1973).
2. Gattner, H.-G. & Schmitt, E. W. *Z. physiol. Chem.* **358**, 105–113 (1977).
3. Inouye, K. *et al. J. Am. chem. Soc.* **101**, 751–752 (1979).
4. Gattner, H.-G. *et al. Z. physiol. Chem.* **361**, 1135–1138 (1980).
5. Tager, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3181–3185 (1980).
6. Geiger, R. *et al. Chem. Ber.* **108**, 2758–2763 (1975).
7. Cao, Q. P. *et al. Scientia sin.* **23**, 1309–1315 (1980).
8. Homandberg, G. A. *et al. Biochemistry* **17**, 5220–5227 (1978).

Comparative antiviral efficiency of leukocyte and bacterially produced human α-interferon in rhesus monkeys

Huub Schellekens*, Arie de Reus*, Reinder Bolhuis†, Michael Fountoulakis‡, Catherine Schein‡, Josef Ecsödi‡, Shigekazu Nagata‡ & Charles Weissmann‡

* Primate Center TNO, PO Box 5815, 2280 HV Rijswijk, The Netherlands
† Rotterdam Radiotherapeutic Institute, 3075 EA Rotterdam, The Netherlands
‡ Institut für Molekularbiologie I, Universität Zürich, 8093 Zürich, Switzerland

Interferons are being tested as antiviral and antitumour agents in man[1–5], but because of their limited availability, clinical trials have remained inconclusive. Most of the interferon (IFN) preparations tested in man have been derived from human peripheral blood, buffy-coat leukocytes (HuIFN-α)[6] induced by Sendai virus or from human diploid fibroblastoid cells (HuIFN-β) superinduced with poly(rI)-poly(rC)[7]. As production of interferon from leukocytes is restricted by the availability of blood donors, and production from cultured cells is a laborious and relatively costly process, the use of recombinant DNA technology may provide the most economical approach to the large-scale production of pure species of human interferon. The testing of such interferons is in its infancy. With the availability of bacterially produced HuIFN-α_2 we have been able to compare its activity against vaccinia virus infection of the rhesus monkey with that of HuIFN-α from leukocyte buffy coats. Here we provide evidence that the two human α-interferons have comparable antiviral activity but that the bacterially produced form has fewer side effects, at least in the rhesus monkey.

There have been reports of the isolation of cDNAs encoding different HuIFN-α species, and of the production of polypeptides with the antigenic and biological properties *in vitro* of interferon from *Escherichia coli* harbouring appropriate hybrid plasmids[8–10]. An important question is whether interferons produced by *E. coli* have the same efficacy *in vivo* as natural interferons. The polypeptides produced in *E. coli* may differ from those produced in leukocytes due to postsynthetic modifications such as proteolytic cleavage, acylation or glycosylation. Recent evidence suggests that, contrary to previous indications, most if not all HuIFN-α species from leukocytes or lymphoblastoid cells are not glycosylated[11], therefore, at least in this regard, we expect interferons made in *E. coli* to be no different. In addition, HuIFN-α from buffy coats is a mixture of several distinct species[11,12] which are encoded by different genes[13]. The different IFN-α species differ substantially in their amino acid sequences (refs 10, 14, 15 and K. Henco, J. Fujisawa and J. Schmid, unpublished results) and in some of their biological activities, determined *in vitro*[14]. In view of these considerations, the effect *in vivo* of a single species of α-interferon produced in *E. coli* may differ from that of a preparation of natural α-interferon. One report has described the protection of

0028–0836/81/340775—02$01.00

© 1981 Macmillan Journals Ltd

squirrel monkeys against a lethal dose of encephalomyocarditis by both natural HuIFN-α and a HuIFN-α from *E. coli*[10]; in this experiment only one very high dosage of the interferons was tested, thus no conclusions could be drawn regarding their relative effectiveness.

There have been reports of rhesus monkeys infected intradermally with vaccinia virus to test the antiviral activity of human interferon[16–18]. The parameters influencing the activities of human interferon *in vivo* were extensively tested in this model system. The pharmacokinetics and the sensitivity of the rhesus monkey to the antiviral activity and side effects of human interferons were similar to those of man. Suboptimal doses of interferon can also be tested in this model.

We have compared the antiviral effects of different doses of HuIFN-α from buffy coats with HuIFN-α from *E. coli* in our rhesus monkey vaccinia model and studied the possible side effects of these products. The natural HuIFN-α, prepared as previously described[6], had a specific activity of $10^{6.2}$ U per mg. HuIFN-α_2 was expressed in *E. coli* from an IFN-α_2 cDNA from which the signal sequence[14] had been removed and replaced by a methionine codon (M. Mishina, W. Boll and A. Hall, unpublished results). Purification of interferon from *E. coli* extracts involved acid precipitation, chromatography on agarose-linked Cibacrone blue and reversed-phase HPLC; the final product had a specific activity of $10^{7.6}$ U per mg protein and showed a single major interferon band and one minor band on SDS-polyacrylamide gel electrophoresis.

Three doses of each interferon (50,000, 150,000 and 500,000 U per kg body weight) were administered to groups of three monkeys each, by daily intramuscular injections, starting the day before challenge with vaccinia and continuing for 7 days after challenge. Three hours after the second interferon treatment the levels of circulating interferon were similar for both types, at all dosage levels (Table 1). All three dosages of natural and bacterial interferon had significant antiviral activity and the degree of protection was dose dependent (Table 1). The highest doses used completely abolished vaccinia-specific lesions. The *in vivo* antiviral activities of the two preparations were similar to their activities *in vitro*. The side effects of the two interferons, however, differed significantly (Table 2). Natural α-interferon from buffy coats, at a dose of 500,000 U per kg, caused leukopenia and fever, whereas the interferon from *E. coli* did not. Liver and kidney function tests showed no adverse effects of either type of interferon. Subsequently, doses of 10^7 U *E. coli* IFNα_2 per kg were administered to three monkeys, again without giving rise to fever, leukopenia, anaphylactic reactions or other evidence of toxicity. The difference in toxic effects between the two interferon preparations when administered to rhesus monkeys could be due to the higher degree of purity of the bacterial product. However, recent work by Scott *et al.*[19] shows that even highly purified leukocyte interferon (NK2-IF), prepared using antibody affinity chromatography, shows the same toxic effect as the much less pure PIF preparation. Perhaps the toxic effects are due to one or more of the α-interferon species other than IFN-α_2, present in the preparation derived from buffy coats. Alternatively, as Genentech's α-interferon (type A, which is equivalent to our α_2) purified from *E. coli* apparently still causes fever and leukopenia in humans (see *Nature* **291,** 105–106; 1981), it could be that the rhesus monkey is not sensitive to the pyrogenic activity of pure α-interferon, and the response we found with the natural interferon was due to impurities in that preparation.

Our study illustrates the use of the vaccinia-infected rhesus monkey as a model for comparing the effects of different interferon preparations. We show that an interferon species produced by recombinant DNA techniques has the same antiviral efficacy *in vivo* as a natural interferon preparation. If the lack of side effects demonstrated in the rhesus monkey proves to be the same for humans, it may be possible to increase the dosage by one order of magnitude or more over that now used.

We thank Dr Kari Cantell for his gift of buffy-coat HuIFN-α and Miss Ellen de Vreede and Mr Hilko D. Wiersema and his assistants for technical assistance. The work at the University of Zürich was supported by the Kanton of Zürich and by a grant from Biogen.

Table 1 Comparison of the efficacy of natural HuIFN-α (buffy coat) and HuIFN-α_2 (*E. coli*) in vaccinia-infected rhesus monkeys

Dose (U per kg body weight)	Circulating interferon HuIFN-α (buffy coat)	Circulating interferon HuIFN-α_2 (*E. coli*)	Lesion score (day 7) HuIFN-α (buffy coat)	Lesion score (day 7) HuIFN-α_2 (*E. coli*)
500,000	480 (±0)	380 (±92)	0.7* (±0.8)	1.2* (±1.0)
150,000	80 (±35)	120 (±43)	1.5* (±0.5)	1.5* (±0.5)
50,000	40 (±17)	73 (±42)	2.2* (±0.8)	2.4* (±1.2)
Untreated controls	<20		4.0 (0.0)	

Rhesus monkeys (*Macaca mulatta*; bred at the Primate Centre, The Netherlands) weighing 1.5–2 kg were used. Natural HuIFN-α was induced in buffy coats using Sendai virus and purified to the stage PIF-1 (see ref. 6); the specific activity was $10^{6.2}$ U per mg protein. HuIFN-α_2 (*E. coli*) was prepared from *E. coli* transformed by the plasmid Z-pBR(Δ Amp)/IFN-α_2(Δ sig)-M2Ib (M. Mishina, W. Boll and A. Hall, unpublished results) and had a specific activity of $10^{7.6}$ U per mg protein. The monkeys were injected intramuscularly with the doses indicated from day −1 until day +7. Each treated group consisted of three monkeys. Four untreated monkeys served as controls. On day 0, the animals were sedated with ketamine hydrochloride, blood was taken and all monkeys were vaccinated with vaccinia virus (Elstree strain, 10^8 plaque-forming units per ml) using a Sterneedle device. At day +7, the skin lesions were scored according to an arbitrary scale[16] from 0 to 4 based on the appearance of papules and pustules, and severity. Results are the mean ±s.d. Interferon activity was assayed using cytopathic inhibition effect test on WISH cells.

* Significant protection at the $P < 0.01$ level when compared with untreated controls in the Mann–Whitney *U*-test. A score of 0.5–1.0 is equivalent to the response found after injection of 0.9% NaCl (refs 16, 17).

Table 2 Comparison of the side effects of HuIFN-α (buffy coat) and HuIFN-α_2 (*E. coli*)

	Leukocyte count ($\times 10^9 l^{-1}$) Day −2	Leukocyte count ($\times 10^9 l^{-1}$) Day 0, 3 h after interferon	Temperature °C (Day 0, 3 h after interferon)
Natural HuIFN-α (buffy coat) (500,000 U per kg body weight)	14.1 (±3.6)	7.2* (±1.3)	38.1† (±1.5)
Bacterial HuIFN-α_2 (500,000 U per kg body weight)	12.6 (±2.7)	15.0 (±3.1)	36.3 (±0.6)
Untreated controls	15.7 (±4.8)	14.0 (±4.1)	36.7 (±0.4)

For experimental details see Table 1 legend. On day −2 and day 0 (3 h after interferon administration), the animals were sedated using ketamine hydrochloride and blood was taken for routine blood chemistry and haematology. At day 0, the body temperature was measured rectally and the monkeys were vaccinated (see Table 1 legend). Values are mean leukocyte counts and temperature ±s.d.

* Significant reduction ($P < 0.05$) when compared with day −2 in the Mann–Whitney *U*-test.

† Significant temperature rise at a $P = 0.053$ level when compared with untreated controls and $P < 0.05$ when compared with the animals treated with HuIFN-α_2 (*E. coli*).

Received 18 May; accepted 27 June 1981.

1. Stewart, W. E. II *The Interferon System* (Springer, New York, 1979).
2. Scott, G. M. & Tyrell, D. A. J. *Br. med. J.* **280,** 1558–1562 (1980).
3. Krim, M. *Blood* **55,** 875–884 (1980).
4. Sikora, K. *Br. med. J.* **281,** 855–858 (1980).
5. Priestman, T. J. *Lancet* **ii,** 113–118 (1980).
6. Cantell, K. & Hirvonen, S. *J. gen. Virol.* **39,** 541–543 (1978).
7. Billiau, A. *et al. Antimicrob. Ag. Chemother.* **16,** 49–55 (1979).
8. Nagata, S. *et al. Nature* **284,** 316–320 (1980).
9. Masucci, M. G. *et al. Science* **209,** 1431–1435 (1980).
10. Goeddel, D. *et al. Nature* **287,** 411–417 (1980).
11. Allen, G. & Fantes, K. H. *Nature* **287,** 408–411 (1980).
12. Rubinstein, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **76,** 640–644 (1979).
13. Nagata, S., Brack C., Henco, K., Schambóck, A. & Weissmann, C. *J. Interferon Res.* **1,** 333–336 (1981).
14. Streuli, M., Nagata, S. & Weissmann, C. *Science* **209,** 1343–1347 (1980).
15. Mantei, N. *et al. Gene* **10,** 1–10 (1980).
16. Weimar, W., Stitz, L., Billiau, A., Cantell, K. & Schellekens, H. *J. gen. Virol.* **48,** 25–30 (1980).
17. Schellekens, H., Weimar, W., Cantell, K. & Stitz, L. *Nature* **278,** 742 (1979).
18. Schellekens, H. & Weimar, W. thesis, Erasmus Univ., Rotterdam (1980).
19. Scott, G. M. *et al. Br. med. J.* **282,** 1345–1348 (1981).

© 1981 Nature Publishing Group

Interaction of the antihypertensive drug felodipine with calmodulin

Stig-Lennart Boström*, Bengt Ljung*, Sven Mårdh†, Sture Forsen‡ & Eva Thulin‡

* AB Hässle Research Laboratories, S-431 83 Mölndal, Sweden and Department of Physiology, University of Göteborg, Sweden
† Institute of Medical and Physiological Chemistry, Biomedical Center, Box 575, Uppsala University, S-751 23 Uppsala, Sweden
‡ Physical Chemistry 2, Chemical Center, POB 740, S-220 07 Lund, Sweden

Despite the marked differences in their chemical structure, members of a heterogeneous group of pharmacological agents are thought specifically to block Ca^{2+} influx through calcium channels[1] and thus mediate negative inotropic cardiac effects and vasodilatation. Electrophysiological studies of the myocardium have shown that the slow inward Ca^{2+} current is blocked[1], but the cellular mechanism of these agents in vascular smooth muscle is largely unknown. Felodipine [4-(2,3-dichlorophenyl)-1,4-dihydropyridine-2,6-dimethyl 3,5-dicarboxylic 3-ethyl-ester and 5-methylester)] is a new antihypertensive agent which seems specifically to dilate precapillary resistance vessels *in vivo*. It is a structural analogue of nifedipine[1] and SKF 24260[2], both of which have been classified as 'calcium antagonists', implying that their vascular as well as myocardial actions are due to a blockade of Ca^{2+} influx. However, the findings reported here point rather to an interaction between felodipine and calcium-binding proteins such as calmodulin.

The effects of felodipine were studied in the isolated rat portal vein. The smooth muscle of this vessel shows many similarities to that of the haemodynamically important small resistance vessels, including its pronounced dependence on extracellular $[Ca^{2+}]$ (refs 3–5). Figure 1*a* shows the spontaneous phasic electrical and mechanical activity of the rat portal vein in the absence of inhibitors. The addition of felodipine (30 nM) produced a gradual decrease of integrated electrical membrane discharge and reduced the amplitude and duration of the contractions (Fig. 1*c*, *d*), whereas tetrodotoxin, at a dose level capable of blocking sodium channels, did not affect this spontaneous activity (Fig. 1*b*).

Figure 2 shows the dose response of the portal vein to cumulative doses of $CaCl_2$ in a K^+-rich solution. The presence of felodipine at low concentration (0.1 pM) in a Ca^{2+}-free solution for 1 h markedly depressed the contractile responses when Ca^{2+} was reintroduced (Fig. 2).

Similar experiments investigated transmembrane Ca^{2+} influx in K^+-depolarized muscle. The Ca^{2+}-depleted portal veins were

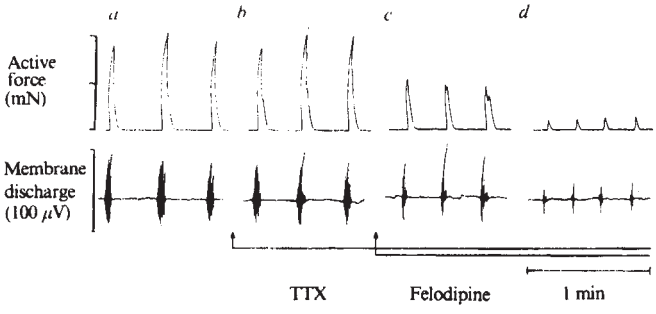


Fig. 1 Effects of felodipine (30 nM) on extracellularly recorded electrical membrane discharge and contractile force of the spontaneously active smooth muscle of the rat portal vein. Vessel mounted for recording of activity in longitudinal media layer[20]. Intermittent bursts of electrical membrane discharge and associated phasic contracting are seen in the control (*a*). Administration of tetrodotoxin (TTX, 1 μM; *b*) in a concentration which would abolish neurogenic vasoconstrictor responses[21], reflecting blockade of Na^+ channels, did not affect the activity. However, addition of felodipine (30 nM) led to a gradual inhibition of membrane discharge and of mechanical force illustrated 20 min (*c*) and 40 min (*d*) after administration.

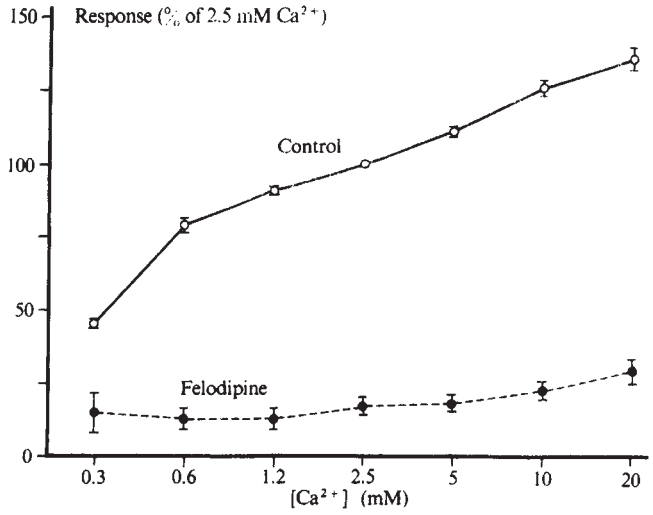


Fig. 2 Effects of felodipine (0.1 pM) on isometric force development in response to graded concentrations of Ca^{2+} in isolated rat portal vein lying in high-K^+ (128 mM) Tris-buffered Krebs solution. Ca^{2+} added to solution after 1 h incubation in nominally Ca^{2+}-free Krebs ('Control') and after a subsequent 1 h in Ca^{2+}-free solution containing the drug. All responses expressed as a percentage of the initial response to 2.5 mM Ca^{2+}. Values obtained in the presence of felodipine were corrected for the decline observed in control experiments where no drug was added (mean ± s.e., $n = 8$). Note the pronounced felodipine blockade of Ca^{2+} responses in the depolarized vascular smooth muscle.

'pulse-labelled'[6] with $^{45}Ca^{2+}$ in 1.2 mM $CaCl_2$ for 2 min, sufficient time for a contracture to develop. The tissues were then washed in a Ca^{2+}-free lanthanum buffer for 1 h to displace extracellularly bound Ca^{2+}. As Fig. 3 shows, the $^{45}Ca^{2+}$ uptake was identical after 1 h incubation with or without 1 nM felodipine. However, after treatment with 2 mM $LaCl_3$ for 1 h, the cellular $^{45}Ca^{2+}$ uptake was <10% of control (Fig. 3). In agreement with these results, Church and Zsoter[7] were unable to detect any alteration in $^{45}Ca^{2+}$ uptake on treatment with nifedipine, verapamil or diltiazem in concentrations producing a clear-cut inhibition of smooth muscle activity. It thus seems that felodipine and some other calcium antagonists interfere with intracellular Ca^{2+} utilization rather than with transmembrane Ca^{2+} influx during activation.

To explore possible interactions between felodipine and myofibrillar proteins, ^{14}C-felodipine was incubated for 1 h with a Ca^{2+}-sensitive crude actomyosin preparation from porcine aorta[8]; samples were then subjected to SDS-polyacrylamide gel electrophoresis. One gel was stained for protein with Coomassie brilliant blue and in the other the distribution of radioactivity was measured in 5-mm sections of the gel. Molecular weight (M_r) estimations were made with standard proteins. Radioactivity was detected only at the dye front and in the 15,000–20,000 M_r region.

Calmodulin (M_r 16,700) occurs in the cytoplasm and is bound to a membrane fraction[9] and to other proteins[10]; its interaction with a drug could therefore affect smooth muscle activity[11–13]. When ^{14}C-felodipine was incubated with calmodulin and subjected to electrophoresis, the radioactivity in the gel was found to coincide with the site of protein staining.

On binding calcium, calmodulin is converted from an inactive to an active form which may influence various cellular processes[14,15], possibly also membrane activation[16]. Apparently, calmodulin can bind a maximum of 4 mol Ca^{2+} per mol protein when assayed at high ionic strength[14,15]. Potentiometric titration of smooth muscle calmodulin with $CaCl_2$ [in 20 mM imidazole buffer (*p*H 7.0, 22 ± 1 °C) with 100 mM KCl] resulted in an equivalence point corresponding to an uptake of 4 mol Ca^{2+} per mol protein, as determined using the Gran-plot technique[17]. However, in a concentration of 4 mol felodipine per mol of protein, the binding was reduced to ~2 mol Ca^{2+}.

In NMR studies, ^{113}Cd has a spin $I = 1/2$ magnetic nucleus and the similarity of the ionic radii of Cd (0.0097 nm) and Ca^{2+}

0028–0836/81/340777—02$01.00

© 1981 Macmillan Journals Ltd

(0.0099 nm) makes Cd^{2+} a good substitute for Ca^{2+} in Ca^{2+} binding protein[18]. The addition of felodipine to $(Cd)_4$–calmodulin produced dramatic changes in the ^{113}Cd-NMR spectrum (Fig. 4), which can be taken to indicate that the binding of felodipine to calmodulin causes the conformation of the latter to change in such a way that at least one of the Ca^{2+}-binding sites is altered. This interaction is apparent at a molar ratio of felodipine to calmodulin of only 0.5 : 1. The ^{113}Cd-NMR results obtained also show that 1–2 mol felodipine are bound per mol calmodulin with a binding constant of 10^5–10^6 mol^{-1}.

Trifluoroperazine, which is known to interact with calmodulin did not inhibit smooth muscle in the rat portal vein (results not shown). Furthermore, their different ^{113}Cd-NMR spectra[18] indicate that the conformational change of calmodulin induced by trifluoroperazine is qualitatively different from that caused by felodipine.

Overall, these results indicate that felodipine may exert its vasodilator effects by interacting with a Ca^{2+}-binding protein rather than by inhibiting Ca^{2+} influx. Church and Zsoter[7] have recently shown that nifedipine and other so-called Ca^{2+} antagonists do not impair transmembrane Ca^{2+} transport in either vascular or cardiac tissue. Furthermore, it is debatable whether Ca^{2+} antagonists do indeed act by selective inhibition of slow Ca^{2+} channels in the myocardium[19]. We suggest that felodipine and possibly other Ca^{2+} antagonists exert at least some of their effects by an action on calmodulin or other Ca^{2+}-binding proteins.

This work was supported in part by grant 3884 from the Swedish MRC.

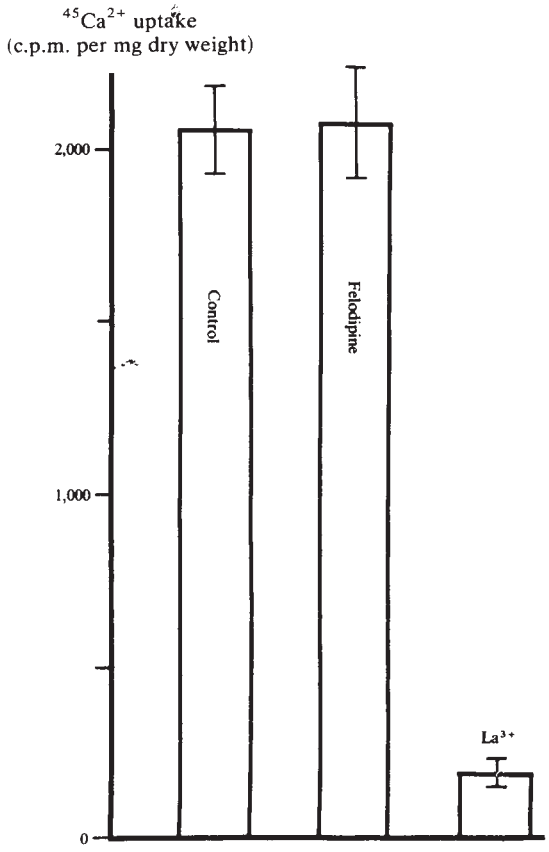


Fig. 3 Effect of felodipine and lanthanum (La^{3+}) on $^{45}Ca^{2+}$ influx in the rat portal vein. After an initial period of 1 h in Tris-buffered Krebs solution, the muscles were kept for 1 h in a Ca^{2+}-free solution. The tissues were then transferred to a high-K^+ (128 mM), Ca^{2+}-free Tris-buffered solution for 10 min. The muscles were then pulse-labelled[6] with $^{45}Ca^{2+}$ for 2 min using $^{45}Ca^{2+}$ in 1.2 mM $CaCl_2$ in the high-K^+ buffer. During such a 2-min $CaCl_2$ exposure a contracture developed in the control specimen. After the initial period experimental muscles were exposed to felodipine (1 nM) or La^{3+} (2 mM). After the 2-min labelling period, the muscles were washed for 1 h in nominally Ca^{2+}-free Tris–Krebs solution including 2 mM $LaCl_2$. The muscles were then blotted lightly, dried overnight at 100 °C, weighed, dissolved and analysed for ^{45}Ca.

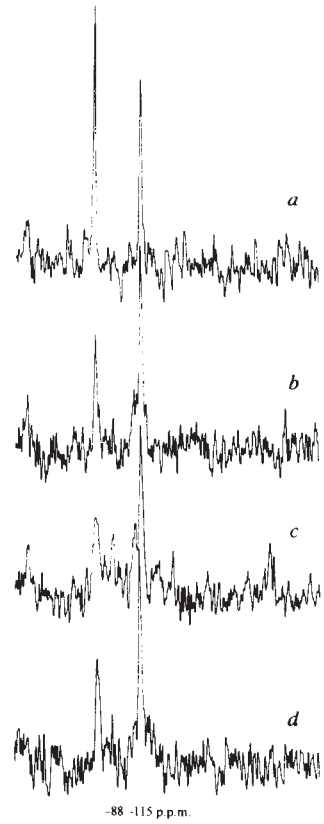


Fig. 4 The effects of felodipine on the ^{113}Cd-NMR spectrum of a 1 mM solution of $(Cd)_4$–calmodulin (bovine testes) at *p*H 8.4 (Tris-perchlorate buffer). In the absence of felodipine (*a*) the two ^{113}Cd-NMR signals are observed at −88 and −115 p.p.m., respectively, relative to 0.05 m $Cd(ClO_4)_2$; higher fields have negative shifts. The molar ratio of felodipine to calmodulin is: *b*, 0.6 : 1; *c*, 0.9 : 1, *d*, 2.0 : 1. The felodipine was added as a 10 mM solution in ethanol–cremaphor RH410 (9 : 1). Apart from a small change in chemical shift of signal at −88 p.p.m. (~2 p.p.m. towards higher fields, the effects of the felodipine solvent corrected for), a broadening of both ^{113}Cd-NMR signals is observed. The broadening which is most marked on the signal at −88 p.p.m. has its maximum at a molar ratio of felodipine to calmodulin of ~1 : 1. Further addition of felodipine causes the signals to become narrower again. The observed effect of felodipine on the ^{113}Cd-NMR spectra of $(CD)_4$–calmodulin is similar to the effect observed on the addition of 1 mol of trifluoperazine per mol calmodulin. The ^{113}Cd-NMR spectra were obtained at 55.54 MHz on a 6.0 T FT NMR spectrometer. The sample volume was 2.5 ml and the sample was placed in a solenoid perpendicular to the axis of the superconducting magnet. Each spectrum is the result of 25,000 transients obtained during 4 h. The sample temperature was 23 ± 1 °C.

Received 23 February; accepted 1 July 1981.

1. Fleckenstein, A. *A. Rev. Pharmac. Tox.* **17,** 149–166 (1977).
2. Taira, N., Himori, N. & Iami, Y. *Clin. exp. pharmac. Physiol.* **3,** 567–574 (1976).
3. Sigurdsson, S. B., Uvelius, B. & Johansson, B. *Acta physiol. scand.* **95,** 263–269 (1975).
4. Sutter, M. C., Hallbäck, M., Jones, J. V. & Folkow, B. *Acta physiol. scand.* **99,** 166–172 (1977).
5. Sutter, M. C. & Ljung, B. *Acta physiol. scand.* **99,** 484–495 (1977).
6. Kroeger, E. A., Marshall, J. M. & Bianchi, C. P. *J. Pharmac. exp. Ther.* **193,** 309–316 (1975).
7. Church, J. & Zsoter, T. T. *Can. J. Physiol. Pharmac.* **58,** 254–264 (1980).
8. Sobieszek, A. & Bremel, R. D. *Eur. J. Biochem.* **55,** 49–60 (1975).
9. Kakiuchi, S. *et al. Adv. Cyclic Nucleotide Res.* **9,** 253–269 (1978).
10. Grand, R. J. A. & Perry, S. V. *Biochem. J.* **183,** 285–295 (1979).
11. Dabrowsky, R., Aromatorio, D., Sherry, J. M. F. & Hartshorne, D. J. *Biochem biophys. Res. Commun.* **78,** 1263–1272 (1977).
12. Hathaway, D. R. & Adelstein, R. S. *Proc. natn. Acad. Sci. U.S.A.* **76,** 1653–1657 (1979).
13. Walsh, M. P., Vallet, B., Autric, F. & Demaille, J. G. *J. biol. Chem.* **254,** 12136–12144 (1979).
14. Cheung, W. Y. *Science* **207,** 19–27 (1980).
15. Wang, J. H. & Waisman, D. M. *Curr. Topics cell. Regulation* **15,** 47–107 (1979).
16. Thorens, S. *J. Muscle Res. Cell Motility* **1,** 455–456 (1980).
17. Dyrssen, D., Jagner, D. & Wengelin, F. *Computer Calculation of Ionic Equilibria and Titration Procedures*, 204–238 (Almqvist & Wiskell, Stockholm; Wiley, New York; 1968).
18. Forsén, S., Thulin, E. Drakenberg, T., Krebs, J. & Seamon, K. *FEBs Lett.* **117,** 189–194 (1980).
19. Vaughan Williams, E. M. *Anti-Arrhythmic Action and the Puzzle of Perhexiline*, 32–33 (Academic, London, 1980).
20. Ljung, B. & Stage, D. *Acta physiol. scand.* **80,** 131–141 (1970).
21. Ljung, B. & Stage, D. *Acta physiol. scand.* **94,** 112–127 (1975).

©1981Nature Publishing Group

MATTERS ARISING

Disruption of meteoritic iron parent bodies

THE discovery by Clarke *et al.*[1] of pre-terrestrial shock polymorphism (ε structure) and shock-induced diamond in Allan Hills A77283 has implications for the process by which the parent body of this meteorite disrupted.

The internal metallographic structures of meteoritic iron seem to require well insulated parent bodies of 10–100 km size[2,3], whereas iron meteorites typically fall to Earth as bodies of 10–100 cm with occasional crater-forming masses of perhaps up to 100 m size. However, of the 70 or so IA irons (excluding A77283) only 3 show sufficient signs of mechanical damage to produce the ε structure[4]. Of these three Cañon Diablo is well known to contain diamond, but is a crater-forming mass and the indications are that its shock effects arose during Earth impact. Cranbourne and Magura are showers, not associated with known craters, and are badly corroded. Magura has been reported to contain diamond[4].

In the absence of ablative heat effects it is difficult to say whether the shock effects in Cranbourne and Magura are pre-terrestrial or not, but the Allan Hills A77283 observations now open up the possibility that they may be.

Thus, the disruption of the IAB parent body seems to have been effected with major damage (ε) to only two or three of the resulting fragments. By contrast, most of the 130 or so members of the IIIAB group show ε structures or shock heating effects.

It is therefore easy to accept a collision process for the disruption of the IIIAB parent body but the marked absence of shock polymorphism in the IAB irons is puzzling. One possibility is that the distribution of non-metal phases (silicates, sulphides) was different in the two parent bodies and allowed damage to be more concentrated in the non-metal portion of the IAB parent. The new evidence on A77283 and the possible pre-terrestrial character of shock effects in Cranbourne and/or Magura now indicate that collision was involved in the disruption of the IAB parent body. The nature of the assumed non-metal portion of the IAB parent which bore the brunt of collision damage remains unresolved but the position of the shocked diamondiferous ureilites might be reconsidered in this context.

H. J. AXON

Metallurgy Department,
University/UMIST,
Manchester M1 7HS, UK

1. Clarke, R. S. Jr, Appleman, D. E. & Ross, D. R. *Nature* **291**, 396–398 (1981).
2. Wood, J. A. in *Asteroids* (ed. Gehrels, T.) 849–891 (University of Arizona Press, 1979).
3. Goldstein, J. I. & Axon, H. J. *Naturwissenschaften* **60**, 313–321 (1973).
4. Buchwald, V. F. *Handbook of Iron Meteorites* (University of California Press, 1976).

Fish versus zooplankton predation in lakes

FISH predation has traditionally been viewed as one of the most important factors regulating zooplankton community structure in freshwater lakes[1,2]. Recently Lane[3], approaching the problem from a different point of view, has conducted comparative studies of vertebrate and invertebrate predation in a temperature freshwater lake with numerous species of zooplanktivorous fish and concluded that piscine predators "probably have little effect on the myriad of interactions among most zooplankton species". Lane's comparative approach for studying predation in aquatic systems is commendable. However, we question her conclusion that fish were unimportant in affecting the zooplankton community which she studied because the following methodological problems of her work would result in a gross underestimate of the relative impact of fishes.

(1) Fish biomasses were poorly estimated. Estimates of absolute densities of fish cannot be made with gill nets unless ancillary studies are conducted (that is, mark and recapture). Additionally, gill nets are not available in mesh sizes which will catch larval and small juvenile fish. Consequently, Lane's estimates of fish density ignored juveniles of the 70 fish species in the lake. Most juvenile fish are planktivorous[4], and although small, are more abundant than large fish; their relative consumption rates are also greater[5]. Therefore, they may significantly affect total piscine predation rates. Only predation by one fish, *Osmerus mordax*, was considered in her study. Consumption by other species was ignored because they were seldom caught in gill nets or were considered "inshore species". However, gill nets are highly selective for some species[6], and absence from these nets does not indicate absence from a system. Inshore species may also have access to pelagic plankton by undertaking diel migrations to the pelagic zone[7,8] or by locating plankton concentrated near the littoral by currents[9]. We have observed both types of behaviour in a zooplanktivorous fish, *Menidia audens* (W.W. and H.L., unpublished results).

(2) Prey consumption rates by fish were severely underestimated. Samples to estimate gut fullness and evacuation rates were taken at only three periods during 24-h cycles: significantly, none was taken during daylight periods. Consequently, feeding peaks may well have been missed. For example, our work with *Menidia* indicates that dawn and dusk feeding peaks would be largely missed with Lane's sampling regime and would result in an underestimate of consumption rates by at least 100%. Lane's data suggest that at least some feeding by *Osmerus* occurs during daylight, as on one of three sample dates, fish were 60% full at sunset. She also assumed a linear instead of an exponential model of gut evacuation and that feeding and gut evacuation are mutually exclusive events. These assumptions, coupled with infrequent sampling, can also lead to severe underestimates of consumption rates[10].

(3) Additionally, there are apparent difficulties in estimating invertebrate predation rates. Lane's previously published estimates of zooplankton predation rates[11,12] differ by more than one order of magnitude. Such differences could lead to considerable errors in interpreting the relative importance of vertebrate and invertebrate predation. We also question the statement that "*Cyclops* . . . perhaps selects prey more for their availability than for any other factor",[3] because prey morphology, size and behaviour contribute significantly to prey selection[13–15].

Lane's work has shown that invertebrate predators must exert an important evolutionary pressure on zooplankton community structure. Additionally, she has emphasized that the effects of fish on the zooplankton may often be indirect[14,16]. However, given the methodological difficulties listed above, we are unconvinced that fish have little effect, whether direct or indirect, on zooplankton communities.

WAYNE WURTSBAUGH

Department of Wildlife and Fisheries,
University of California,
Davis, California 95616, USA

HIRAM LI
JUDITH LI

Oregon Cooperative Fishery Research Unit,
Department of Fisheries and Wildlife,
Oregon State University,
Corvallis, Oregon 97331, USA

1. Hall, D. J., Threlkeld, S. T., Burns, C. W. & Crowley, P. H. *A. Rev. Ecol. Syst.* **7**, 177–208 (1976).
2. O'Brien, W. J. *Am. Scient.* **67**, 572–581 (1979).
3. Lane, P. A. *Nature* **280**, 391–393 (1979).
4. Nikolsky, G. V. *The Ecology of Fishes* (Academic, New York, 1963).
5. Brett, J. R. in *Fish Physiology* Vol. 8 (eds Hoar, W. S. *et al.*) 599–667 (Academic, New York, 1979).

0028–0836/81/340779—02$01.00

© 1981 Macmillan Journals Ltd

6. Lagler, K. F. in *Fish Production in Fresh Waters* (ed. Bagnal, T.) 7-47 (Blackwell, New York, 1978).
7. Hall, D. J. *et al. J. Fish. Res. Bd Can.* **36**, 1029-1039 (1979).
8. Bohl, E. *Oecologia* **44**, 368-375 (1980).
9. George, D. D. & Edwards, R. W. *J. appl. Ecol.* **13**, 667-690 (1976).
10. Elliott, J. M. & Persson, L. *J. Anim. Ecol.* **47**, 977-991 (1978).
11. Lane, P. A., Klug, M. J., & Loudon, L. *Trans. Am. Microsc. Soc.* **95**, 143-155 (1976).
12. Lane, P. A. *Verh. int. Verein. theor. angew. Limnol.* **20**, 480-485 (1978).
13. Li, J. L. & Li, H. W. *Limnol. Oceanogr.* **24**, 613-626 (1979).
14. Kerfoot, W. C. *Limnol. Oceanogr.* **22**, 316-325 (1977).
15. Gilbert, J. J. & Williamson, E. E. *Oecologia* **37**, 13-22 (1978).
16. Northcote, T. G., Walters, C. J. & Hume, J. M. B. *Verh. int. Verein. theor. angew. Limnol.* **20**, 2003-2012 (1978).

LANE REPLIES—Wurtsbaugh *et al.* state that I concluded fish to be unimportant with regard to zooplankton community structure[1]. I did not draw this conclusion. They confuse two questions: (1) Which predators, vertebrate or invertebrate, exert the most predation pressure on lake zooplankton?, and (2) what is the significance (importance) of these predation pressures? My paper was concerned with providing a quantitative answer for question (1). With regard to question (2) I was careful not to make value judgements. I stated that vertebrate predators often have "dramatic effects" on lake zooplankton. I shall reply to the three methodological criticisms of Wurtsbaugh *et al.*, then clarify the problem of significance (silver bullet obsession).

(1) Fish density: my report gave only a small part of a 4-yr collection effort of the Gull Lake fish community. In addition to gill nets, mark and recapture, a high-speed sampler, a purse seine, sonar, dip nets and hand lines were used for field sampling, and experimental cages were used for predation studies on immature fish. We did not have access to a midwater trawl which has been successfully used in the Great Lakes[2]. Our laboratory studies showed that clipped smelt were highly susceptible to fungus infections and exhibited high mortalities. Others have found summer marking to be impractical[3]. Consequently, we abandoned this method of estimating fish populations. As Wurtsbaugh *et al.* point out, most types of fishing gear are selective[4]; however, sonar traces gave reasonable agreement with gill-net results.

They are correct in stating that juvenile fish "may affect significantly total piscine predation rates" for some environments and there is undoubtedly some error there. Their argument for designating juvenile fish as important predators was based on the reasoning that they are more numerous and have greater relative consumption rates than adults. When I used a similar argument in comparing vertebrate with invertebrate predators, Wurtsbaugh *et al.* failed to acknowledge it. The reasoning is correct for both arguments, if smaller predators are more numerous. The combined sampling methods failed to reveal large numbers of immature fish occupying the central station. References to littoral *Menidia audens*, the Mississippi silverside, which inhabits shallow, warm Clear Lake, California, are irrelevant to the Gull Lake situation for almost all comparable criteria[5-15]. My study[1] was restricted to the pelagic zone.

Fish consumption rates: all diurnal sampling was done with four periods as Table 1 of my paper clearly shows. Use of sonar to check the absence of fish during the day and their diurnal movements were mentioned in Fig. 1 legend. In a subsequent study, smelt were collected on transects from the central station to shore. Data on their stomach contents provided no evidence of inshore feeding during the day or of a dawn feeding peak. Wurtsbaugh *et al.* make two useful points about linear compared with exponential models of gut evacuation and the assumption of independence of feeding and gut evacuation. Usually gut evacuation time and feeding periodicity are determined experimentally. The smelt did not cooperate in our laboratory studies thus I kept the model as simple as possible.

Invertebrate predation rates: these do differ by an order of magnitude over the large range of conditions in my experiments. Feeding rates for filter-feeding zooplankton are also as variable; density relationships are even more variable. Environmental heterogeneity and feeding periodicity also contribute to rate variations. I stated in Fig. 1 legend[1] that many other invertebrate predators consume the prey species; this would underestimate the degree of invertebrate predation extant in Gull Lake[16,17].

(2) Significance: I concluded that smelt account for ~5% of the total predation on Gull Lake zooplankton[1]. This does not mean that invertebrate predators are 20 times more important than vertebrate predators. Important to whom?—A particular population? The daphnids? The community? Important in what way?—For calculating nutrient and energy budgets? For determining community structure and trophic networks? For quantifying stability?

In aquatic ecology there is an obsession with the silver bullet—the belief that a single factor can be found to explain ecosystem dynamics. For example, the limiting factor in eutrophication studies[18], the arguments on density-dependent versus independent population regulation, the controversy regarding whether competition or predation regulates the intertidal zone and now the pelagic zone, and the suggestions that aquatic ecosystems should be organized around the axis of body size. Some of these are parameter inputs. They do not control the ecosystem but rather set particular variables in motion. Variables included within a network possess similar rates whereas parameters are faster or slower. Wurtsbaugh *et al.* believe that fish are a variable—this may be true for littoral species in particular. In testing the Gull Lake data set with loop analysis, it seems that smelt act as a parameter input to *Chaoborus* spp. and large *Daphnia pulex*. Inclusion of smelt as a variable reduced the agreement of model predictions with the data. In fact, smelt are not even the predominant input—those enter at nutrient and algal levels. This is why I stated that "smelt probably have little effect on the myriad of interactions among most zooplankton species".

It is not possible to take one predation link out of a whole ecological network and declare that this controls the system or is the most important part of the dynamics[18]. The effect of any given parameter or variable on another variable is a consequence of the configuration of the whole network. In addition, indirect pathways often swamp direct ones. Smelt have reduced *Chaoborus* spp. populations, which in turn have eased the predation pressure on many small prey species. Thus, ignoring indirect pathways could cause substantial overestimates of smelt predation. Similar results have been noted elsewhere[19]. This positive effect on small species demonstrates that ecologists should not hasten to load their ecological pistols with silver bullets for they will undoubtedly be caught in the cross-fire.

PATRICIA A. LANE

Department of Population Sciences, Harvard School of Public Health, Harvard University, Boston, Massachusetts 02115, USA
and
Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada

1. Lane, P. A. *Nature* **280**, 391-393 (1979).
2. MacCallum, W. R. & Regier, H. A. *J. Fish. Res. Bd Can.* **27**, 1823-1846 (1970).
3. O'Connor, J. F. & Power, G. *Naturaliste can.* **101**, 755-762 (1974).
4. Hamley, J. M. *J. Fish. Res. Bd Can.* **32**, 1943-1969 (1975).
5. Cook, S. F. & Moore, R. L. *Trans. Am. Fish. Soc.* **99**, 70-73 (1970).
6. Moyle, P. B., Fisher, E. W. & Li, H. W. *Calif. Fish Game* **60**, 144-149 (1974).
7. Hubbs, C., Sharp, H. B. & Schneider, J. F. *Trans. Am. Fish. Soc.* **100**, 603-610 (1971).
8. Mense, J. B. *Bull. Okla. Fish Res. Lab.* **6**, iii-32 (1967).
9. Elston, R. & Bachen, B. *Trans. Am. Fish. Soc.* **105**, 84-88 (1976).
10. Burbridge, R. *Trans. Am. Fish. Soc.* **98**, 631-640 (1969).
11. McKenzie, R. A. *J. Fish. Res. Bd Can.* **15**, 1313-1327 (1958).
12. Ferguson, R. G. *Great Lakes Res. Div.* **13**, 47-60 (1965).
13. Li, H. W., Moyle, P. B. & Garrett, R. L. *Trans. Am. Fish. Soc.* **105**, 404-408 (1976).
14. Goldman, C. R. & Wetzel, R. G. *Ecology* **44**, 283-294 (1963).
15. Moss, B. *Freshwater Biol.* **1**, 309-320 (1972).
16. Lane, P. A. *Verh. int. Verein. theor. angew. Limnol.* **20**, 480-485 (1979).
17. Lane, P. A., Klug, M. J. & Louden, L. *Trans. Am. Microsc. Soc.* **95**, 143-155 (1976).
18. Lane, P. A. & Levins, R. *Limnol. Oceanogr.* **22**, 454-471 (1977).
19. Northcote, T. G., Walter, C. J. & Hume, J. M. B. *Verh. int. Verein. theor. angew. Limnol.* **20**, 2003-2012 (1978).

© 1981 Nature Publishing Group

BOOK REVIEWS

Mechanisms of scientific revolution

G. N. Cantor

A World on Paper: Studies on the Second Scientific Revolution. By Enrico Bellone. Pp.220. ISBN 0-262-02147-1. (MIT Press: 1980.) $20.25, £10.50.

HISTORIANS of science have for a long time been interested in periods of major, rapid and partially discontinuous change in science. The most radical of these changes dates from, roughly, the first half of the seventeenth century and is generally known as *the* scientific revolution. Other more parochial revolutions, such as the Darwinian revolution, have also attracted much attention. In physics two main periods perhaps deserve the appellation "revolution"; one associated with Fresnel's work on light, Carnot's and Fourier's on heat and other developments in the late eighteenth and early nineteenth centuries, and the other related to the rise of quantum and relativity theories a century later. Whether we decide to characterize either of these as a "revolution" depends on how we define the term in respect to changes in many facets of science.

Enrico Bellone's *A World on Paper: Studies on the Second Scientific Revolution* is concerned, as the sub-title suggests, with substantiating the claim that a revolution occurred in physics in the late eighteenth and early nineteenth centuries. The author claims that during this period physics became more mathematical, theoretical and abstracted from experience; physicists thus constructing "a world on paper". This book, first published in Italian in 1976, now appears in English translation. Stillman Drake, an authority on Galileo, has provided a somewhat eulogistic foreword. Bellone's book encompasses an important topic for current historical research, since despite the considerable amount of respectable secondary literature on this period there is comparatively little of a synthesizing nature. Unfortunately, this study is disappointing and fails to achieve its objective.

Problems arise partly because Bellone, a professor of the history of science at the University of Genoa, is anxious to demolish what he sees as two pernicious and pervasive doctrines. First, he intends to refute the view, which he appears to attribute indiscriminately to philosophers and most historians of science, that from the seventeenth to the end of the nineteenth century physics was dominated by the doctrine of mechanism. Not only does he fail to explain what he understands by this term but he also does not name his opponents, with the exception of Pierre Duhem. Yet Duhem's aim was to *contrast* the British tendency to construct mechanical models of physical phenomena with the continental methods of mathematical abstraction. *The Aim and Structure of Physical Theory* was published 75 years ago and no respectable historian or philosopher of science in recent years has championed any simplistic thesis concerning the total dominance of mechanism during the nineteenth century. We can conclude that Bellone has set up a straw man as his major opponent.

The second of Bellone's opponents is less of a straw man than an inappropriate one. Apparently with Kuhn, Popper, Feyerabend and Lakatos in mind, he attacks those philosophers who have proposed general theories of scientific development. However, in order to controvert, for example, Kuhn's theory it is necessary to show that either it is intrinsically incoherent or it does not accord with the actual development of science. Bellone attempts a rather inadequate variant of the latter strategy by arguing in five case-studies that physicists have articulated their own views about scientific method and have not been indebted to philosophers of science, least of all to those who espoused mechanism.

By far the longest study is an attempt to show that, contrary to Duhem's characterization, Kelvin was not a mechanist. To substantiate this claim the author discusses Kelvin's work in such areas as energy conservation and the vortex atom hypothesis. However, he seems prepared to concede that Kelvin deployed mechanistic hypotheses and analogies provided they were adequately grounded on known physical laws and expressed in mathematical terms. More positively he interprets Kelvin as being centrally concerned with inductively discovered physical laws and prepared to employ mathematics only as a useful tool. These attitudes are traced to the views of John Herschel and Thomas Reid. Kelvin's empiricist philosophy is contrasted, in the second study, with that of Boltzmann who propounded a mathematical and highly abstract form of physical theorizing which set him against Mach's sensationalism and Ostwald's theory of energetics, and also led to his work on probability calculus. Of particular importance to Bellone is Boltzmann's partial rejection of mechanistic explanations.

The third study is largely concerned with Ampère who is interpreted as having rejected action-at-a-distance forces — as championed by Newton and Laplace — and adopted a contact-action ether theory of electromagnetism. There follows an essay contrasting the fluid theory of electricity held by the Abbé Nollet in the 1740s with the mathematical instrumentalism propounded by Cavendish a quarter of a century later. Finally we return to the Kelvin–Boltzmann dispute over the respective roles of mathematics and experience and to Kelvin's recognition in 1900 that physics was facing two major problems: first the problem of reconciling ether theory with the Michelson–Morley experiment and, second, the inadequacy of the Maxwell–Boltzmann theorem for the equipartition of energy. These difficulties perceived at the end of the nineteenth century are interpreted by Bellone as indicating continuity with twentieth-century physics; thus, in his opinion, no scientific revolution occurred early this century.

I hope that I have accurately conveyed the author's major claims since I found this a surprisingly difficult, even opaque, book. While admiring Bellone's courage in attempting to cover a major historical topic, I encountered several difficulties with his thesis. Some problems stem from disagreements with his historical claims. To give two examples: there is strong evidence, some of which is cited in this book, that Ampère did not subscribe to a contact-action ether, and Kelvin's philosophy cannot be simply traced back to Reid's — indeed they differed radically over the question of hypotheses and analogies. Other difficulties arise because although the book contains summaries of various writers' highly interesting views on mathematics, experimentation, theory and mechanics, the presentation of these issues lacks coherence and is weak in analysis, so that the reader is frequently assailed by ambiguous statements and points which need both clarification and evidential support.

Yet more problems stem from the historiography which is based on what the author calls the "scientist's dictionary" (considered by Professor Drake to be Bellone's major contribution to the study of history). The dictionary of a scientist contains, it appears, the set of his currently adopted ideas. If I have interpreted Bellone correctly, then his theory of dictionaries is unhelpful if not positively misleading.

© 1981 Nature Publishing Group

There are many significant dissimilarities between a mind and a dictionary: a dictionary, unlike a mind, contains only definitions of words, is open to public scrutiny and so on. How are we to construct a scientist's (private) dictionary from his public statements when the latter are often shaped for the audience for which they are intended? Given these and similar queries it is unfortunate that Bellone did not discuss this theory in much greater detail.

Bellone's notion of the scientist's dictionary has a further consequence for his historical claim that a revolution occurred in physics in the late eighteenth and early nineteenth centuries. Revolutions have, as Kuhn insisted, a social component which he expressed in terms of paradigm change being a change in consensus among the members of a scientific *community*. Even if one does not agree with the details of Kuhn's thesis it would appear difficult for the historian interested in large-scale changes in science to omit this social dimension. Yet this is precisely what Bellone does. According to his theory, each scientist maintains his own dictionary and yet Bellone offers the historian no means of moving from individual dictionaries to an understanding of broader historical processes. Indeed, his failure to appreciate the latter leaves the reader unclear as to what constituted the "second scientific revolution".

Unfortunately, the author conveys the impression that he is divorced from the main currents of historical scholarship. An examination of the book's endnotes reveals surprisingly few references to the major secondary works on the subjects he discusses. His indiscriminate use of terms such as deduction, crucial experiment, common sense, scientific revolution, mechanism and Newtonianism indicates that he is not aware of the specific meaning conventionally attached to them. Finally, it is surely only ignorance of recent scholarship that could have led the author to construct his book as an attack on Duhem's outmoded thesis.

G. N. Cantor is a Lecturer in the Department of Philosophy at the University of Leeds.

Across the spectrum of IR astronomy

R.F. Carswell

Infrared Astronomy. International Astronomical Union Symposium No. 96. Edited by C.G. Wynn-Williams and D.P. Cruikshank. Pp.376. ISBN hbk 90-277-1277-1; ISBN pbk 90-277-1228-X. (Reidel: 1981.) Hbk Dfl.90, $47.50; pbk Dfl.45, $23.50.

FEW branches of astronomy have made such rapid progress in the past few years as the infrared. With the continuing development of modern detector systems, and in particular the availability of four telescopes high on Mauna Kea in Hawaii above most of the atmospheric absorption, there is every expectation that this will continue. It was therefore appropriate that a symposium on infrared astronomy should be held in Hawaii last year, drawing together contributions from many of the most active people in the field.

The papers included in the proceedings reflect the diversity of topics to which infrared astronomy is making its significant, and often unique, contribution. Articles on planetary studies, protostellar objects, molecular clouds, dust, nebular emission lines, stars in external galaxies, and active galactic nuclei show just how wide the range is. The often dull (to an outside astronomer) technical papers describing detector systems were deliberately excluded from the symposium, and the editors chose to include only reviews on each topic rather than all the inevitable specialist results normally heard at such symposia. The result is a volume which summarizes the present state of the subject well and gives an idea of the rapid advances being made without the non-specialist reader having to wade through a mass of detail.

To choose a few at random, the three descriptions by Evans, Hyland and Thompson highlight the value of infrared studies of the highly obscured regions where star formation is occurring, and provide useful summaries for the relatively ignorant such as myself. The review of the reviews on this topic by Zuckerman helps to give the overall picture. In an area where I have more expertise, Soifer and Neugebauer give us a comprehensive survey of the properties of active galactic nuclei and quasars.

The final summary by Longair states succinctly what emerges from the proceedings in general. Infrared astronomy is proving to be useful in probing an increasingly diverse range of astrophysical questions and, with future developments, such as grating spectrometers, area detectors and infrared astronomy satellites, the trend towards "in depth" studies of particular classes of objects at all wavelengths is likely to continue. Thus, a book on infrared astronomy is likely to give way to more specialist collections. In the meantime, this volume provides a useful, up-to-date review of the entire field.

R.F. Carswell is at the Institute of Astronomy, Cambridge.

Hot history

P.T. Landsberg

The Tragicomical History of Thermodynamics 1822–1854. Studies in the History of Mathematics and Physical Sciences, 4. By C. Truesdell. Pp.372. ISBN 3-540-90403-4. (Springer-Verlag: 1981.) DM99, $58.50.

WHAT is the meaning of the following result: "If the total heat of a homogeneous and uniformly hot substance be conceived to be divided into any number of equal parts, the effects of those parts in causing work to be performed are equal"? If this should not be quite clear, the following may help: "If the absolute temperature of any uniformly hot substance be divided into any number of equal parts, the effect of those parts in causing work to be performed are equal". These are quotes from William John Macquorn Rankine (1820–1872), a Scottish mining engineer, writer on thermodynamics, inventor of the word "adiabatic". They are his version of the second law of thermodynamics as given in his *Manual of the Steam Engine* which saw many editions, starting in 1859. Little wonder, then, that James Clerk Maxwell felt in 1878 that anyone who actually understood these remarks could explain thermodynamically what Tennyson had said of the great Duke:

whose eighty winters froze with one rebuke
All great self-seekers trampling on the right.

If oscillations of this order in the clarity of conceptions concerning thermodynamics could take place *after* the period covered in Truesdell's book, it is perhaps not surprising that the period itself also furnishes many challenges to those who come in today to tidy up the errors of the founding fathers of the subject. This is precisely what Truesdell seeks to do (the above quotations will not be found in this book).

Where Maxwell proceeded with delicacy and humour, Truesdell is hard-hitting, hoping to raise in this way the subject from the "*Dismal Swamp of Obscurity* that from the first it was and that to-day in common instruction it is" (p.6). Leaving an assessment of this provocative remark to the many students and teachers of the subject, the book deals with Truesdell's rediscovery (following J. Moutier) of F. Reech (1805–1884) and has chapters or sections on Carnot, Helmholtz (*The Conservation of Force* being described as his weakest paper), Fourier, Clausius, Rankine, Kelvin and many others, mathematical aspects being given prominence.

The teaching of the subject can conveniently start with the introduction of specific and latent heat functions C and l defined in a usual notation by:

$$dQ = C_v\,dT + l_v\,dV = C_p\,dT + l_p\,dp.$$

To these relations a third relation may be

© 1981 Nature Publishing Group

added, which is not independent but yields a pleasant symmetry:

$$dQ = m_v\, dV + m_p\, dp.$$

This procedure has been used in a number of text books, including my own. Truesdell relates it to what he calls the "Doctrine of Latent and Specific Heats".

The book makes interesting reading and in its emphasis on the history of mathematical formulations of thermodynamics it breaks new ground. It suffers by the awkwardness of the chapter arrangements, cross references and notation, which makes it hard to read, and by the very limited references to the writings of contemporary historians of science, which sets it apart from current historical research. Not even Clausius' 1872 paper on the history of the mechanical theory of heat is cited as a source. I conclude with quotations illuminating the harsh ethos of the work:

Of Laplace: As a physicist he preferred contorted structural hypothesis, as a mathematician he was unusually loose, even for his day [p.41].

Of Fourier: I doubt if any critical reader today would find in his book much insight into nature beyond that common in works of his predecessors and contemporaries [p.76].

"This unhappy quality is the tough and tortuous thread of the plot of the tragicomedy. It began with Laplace and was spun out by Fourier" (p.77).

P.T. Landsberg is Professor of Mathematics at the University of Southampton.

Will plant biochemistry go to the ball?

Philip Rubery

The Biochemistry of Plants: A Comprehensive Treatise. Editors-in-chief P.K. Stumpf and E.E. Conn. Vol.1 *The Plant Cell* (edited by N.E. Tolbert), pp.699, ISBN 0-12-675401-2. Vol.2 *Metabolism and Respiration* (edited by David D. Davies), pp.687, ISBN 0-12-675402-0. Vol.3 *Carbohydrates: Structure and Function* (edited by Jack Preiss), pp.638, ISBN 0-12-675403-9. (Academic: 1980.) $65, £43 per volume.

A RECENT issue of *Trends in Biochemical Sciences* carried the headline "Plant biochemistry — a Cinderella too long". Does the publication of a comprehensive treatise on the biochemistry of plants herald the arrival of this poor relation in good society? As the editors comment in their general preface, the study of plants has had a mixed reception in the biochemical community. Thus the treatise aims not only to instruct but also to inspire — indeed, evangelize.

The first volume is largely self-contained and is organized into chapters centring on particular organelles and cellular membrane components which collectively cover many of the topics dealt with under the more system-based approach of Vols 2 and 3. Indeed, from reading all three volumes one theme that constantly recurs is the unity of plant cells and the interdependence of organelles. For example, an overall account of the current status of carbon flow and compartmentalization in photorespiration and gluconeogenesis from fat can be synthesized from articles in Vol.1 on microbodies, on mitochondria and on chloroplasts; in Vol.2 on control of carbohydrate oxidation and on photorespiration; and in Vol.3 on integration of carbohydrate metabolism. The glyoxylate cycle is also covered in Vol.4 and a further chapter on photorespiration is promised for Vol.8.

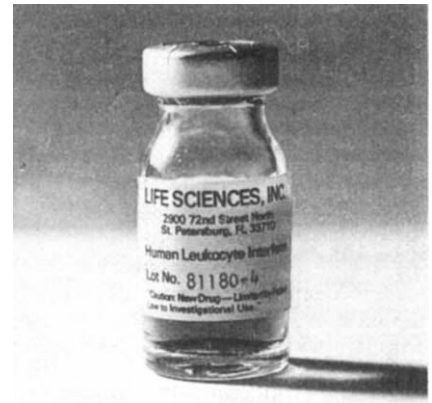
One of the primary goals of a treatise such as this is to provide authoritative accounts of major areas which embody a large corpus of research. Many of the chapters fulfil this aim well — those on sugar nucleotide transformations (Vol.3), polysaccharide solution conformation (Vol.3), development and evolution of plastids (Vol.1) and mitochondrial electron transport and energy coupling (Vol.2) are particularly good. The chapter on primary cell walls (Vol.1) includes a welcome critical review of the evidence for covalent linkages between macromolecules having a key role in wall structure.

A comprehensive account of plant biochemistry has other important tasks to perform. For instance, the applicability to plants of research strategies evolved in other contexts should be assessed and publicized. This is epitomized by ap Rees (Vol.3) in his consideration of the theoretical and experimental approaches to the regulation of hexose phosphate metabolism along the lines influentially described by Newsholme and Start in *Regulation in Metabolism* (Wiley, 1973). Turner and Turner cover similar ground in Vol.2, where Wiskich also deals rather more phenomenologically with the control of the Krebs cycle.

Areas where plant biochemists are

Volumes 4 (*Lipids: Structure and Function*), 5 (*Amino Acids and Derivatives*) and 6 (*Proteins and Nucleic Acids*) in this series have now appeared. The last two volumes making up the *Treatise* (Vol.7 *Secondary Plant Products* and Vol.8 *Photosynthesis*) are scheduled for publication shortly. The completed series will be reviewed in *Nature* at a later date.

LIFE SCIENCES, INC.



Supplying Human Leukocyte Interferon for clinical tests and research.

2900-72nd Street North
St. Petersburg, Florida 33710
(813) 345-9371

Circle No.26 on Reader Enquiry Card.

ATTENTION TECHNICAL INSTRUCTORS!

THE AMERICAN NUCLEAR SOCIETY OFFERS THIS TEXTBOOK SERIES TO THE EDUCATIONAL COMMUNITY:

- *El Wakil*, **NUCLEAR ENERGY CONVERSION**, 682 pages, 6 x 9 hardbound, ANS Order No. 350001 **$28.60 ea.**
- *El Wakil*, **NUCLEAR HEAT TRANSPORT**, 514 pages, 6 x 9 hardbound, ANS Order No. 350002 **$26.80 ea.**
- *Karl Wirtz*, **LECTURES ON FAST REACTORS**, 310 pages, 8½ x 11 softbound, ANS Order No. 350003 **$18.00 ea.**
- *Murray/Cobb*, **CONCEPTS AND CONSEQUENCES**, 710 pages, 7 x 9½ hardbound, ANS Order No. 350004 **$13.95 ea.**

SEND FOR INFORMATION ON HOW TO OBTAIN A DESK COPY.

Please take my order for:
_____Copies of ANS Order No. 350001 $_____
_____Copies of ANS Order No. 350002 $_____
_____Copies of ANS Order No. 350003 $_____
_____Copies of ANS Order No. 350004 $_____
TOTAL $_____

NAME: _____
ADDRESS: _____
CITY: _____
STATE: _____
ZIP: _____

Send to:

AMERICAN NUCLEAR SOCIETY
555 North Kensington Avenue
La Grange Park, IL 60525, USA

Payment enclosed in the amount of $_____
Bill me and I will pay the additional postage and handling charge.

7/81

Circle No.05 on Reader Enquiry Card.

© 1981 Nature Publishing Group

demonstrably ignorant compared to their animal colleagues (the Cinderella effect?) are frequently pointed out and many chapters suggest where plant horizons could be broadened or at least examined more closely. According to Storey (Vol.2), none of the mitochondrial redox flavoproteins has been characterized as an isolated enzyme. The contrast with the wealth of spectroscopic data available for plant mitochondria exemplifies the comparatively patchy sophistication of plant biochemistry. Similarly, the important enzyme thymidylate synthase, discussed by Cossins in Vol.2, has not been characterized from plant tissue. On a wider scale, the biochemistry of many basic physiological problems remains obscure, especially when they lie in the interstices of traditional major domains. For example, we have little detailed understanding of the differences in metabolite traffic, involved in energy supply and carbon skeleton provision, between chloroplasts, cytoplasm, mitochondria and peroxisomes in leaf cells in the light and the dark. Graham's chapter (Vol.2) on the effects of light on "dark" respiration addresses some of these questions. The excellent article on translocation of sucrose and oligosaccharides, by Giaquinta in Vol.3, combines a physiological and biochemical approach to highlight key problems such as the mechanism of sucrose unloading from phloem whose further elucidation could have agronomic and economic implications. It is important that plant biochemistry should, like mammalian biochemistry, be increasingly concerned with applying the burgeoning range of new techniques to essentially cellular problems.

As to omissions, I would have liked some treatment of the cytoskeleton in plants and also perhaps a chapter explicitly devoted to transport across the membranes of organelles. The virtual exclusion of growth substances from the consideration of cell wall dynamics and of seed reserve mobilization is also regrettable, although admittedly phytohormones are not within the plan of the treatise. Nevertheless, most topics that one would hope to find adumbrated by the volume titles are covered although the "metabolic" content of Vol.2 is largely carbohydrate; lipids, amino acids, proteins and nucleic acids, and secondary products being dealt with in Vols 4,5,6 and 7 respectively. Much of the contents of these three volumes and of the treatise as a whole are likely to remain valuable sources of reference and perspective for the rest of the decade. They are not expensive in today's terms for what they provide and the semi-autonomous nature of the volumes should enable a variety of permutations to serve the interests of different groups and individuals. The treatise is timely and, like Cinderella, the subject is of some promise. □

Philip Rubery is a Lecturer in Biochemistry at the University of Cambridge.

Muscle: from hypertension to erection

Andrew P. Somlyo

Smooth Muscle: An Assessment of Current Knowledge. Edited by E. Bülbring *et al.* Pp.576. ISBN 0-7131-4348-7/0-292-77569-5. (Edward Arnold/Texas University Press: 1981.) £45, $95.

SMOOTH muscle cells in blood vessels, bladder, uterus, gastrointestinal tract and airways support and regulate, directly or indirectly, nearly all bodily functions. Nevertheless, physiologists, until recently, have shied away from studying these cells because of their small size, relatively poor ultrastructural organization and great individuality. Our knowledge of the cell physiology of smooth muscle, with a few notable exceptions, is largely due to studies conducted within the past 15 years and, thus, a timely subject for the stated aim of this book — "an assessment of current knowledge". The book is a "family affair"; the editors and contributors are Professor Bülbring of Oxford and her former pupils who, individually and in collaboration, have contributed vigorously to this field. Nevertheless, the volume is not parochial; geographically, the authors range from Oxford to continental Europe, including Russia, and through the United States to Australia and Japan. Intellectually, the contents extend to cover work from laboratories outside "the family" and successfully reconcile previously opposing views, such as the normal occurrence of graded changes in membrane potential in tonic smooth muscles and the existence of smooth muscle regulation by mechanisms independent of the membrane potential (pharmacomechanical coupling).

The major emphasis is on electrophysiology and ion transport, areas of research in which Professor Bülbring and her pupils have been most active. Of what is known about these subjects, the contributions by Brading, Casteels, Jones, Kuriyama and Tomita leave little uncovered. Jones, perhaps having planned successfully his chapter as the basis of a graduate course on smooth muscle, presents a particularly broad view, covering not only relevant aspects of the ultrastructure, electrophysiology and the ionic permeability of normal smooth muscle, but also the membrane properties of vascular smooth muscle altered by high blood pressure. Voltage clamping of smooth muscle is assessed with admirable caution by Bolton, Tomita and Vassort. The application of this technique to tissues having such complex geometries as smooth muscle is a terrain where most researchers fear to tread, and, therefore, the identification of the ionic currents responsible for action potentials in smooth muscles has been ambiguous. The task of identifying such currents is not made easier by the large variations among different smooth muscles in which the contributions of, respectively, sodium and calcium to the early current are quite variable. Burnstock's contribution on smooth muscle development and innervation and Gabella's extensive discussion of the ultrastructure of smooth muscle successfully relate structure to function. Gabella, incidentally, also reflects the consensus of experienced electron microscopists that, in both relaxed and contracted smooth muscle, myosin is organized in thick filaments.

Given the highly individual properties of smooth muscle cells in different organs and the variety of their functions, the subjects to be covered are limited only by space and the authors' interests. These range from the aforementioned studies of vascular smooth muscle altered by hypertension to the role of penile blood vessels in erection. It may come as a surprise to some that erection is due to the relaxation of smooth muscle of the penile arteries, the flaccid state being the result of vasoconstriction.

The flaws of the book are minor, although the mistaken identification of one of the editors, a Professor of some years standing, as an Assistant Professor is certainly an editorial "blooper". There is considerable overlap between the various chapters covering electrophysiology and ion distribution, perhaps unavoidable in a multi-author volume that is the product of a single scientific "school", and also forgiveable in a book that boasts several thousand references dealing with this subject. It is not a textbook, since some important aspects of smooth muscle, including biochemistry, muscle mechanics and energetics did not fall within this group's research interests and are, therefore, discussed little or not at all. In this sense, this book and the recently published volume in the *Handbook of Physiology* series (*Vascular Smooth Muscle*; American Physiological Society, 1981), of which, I should confess, I am an editor, complement each other well; together they provide an encyclopaedic survey of the field.

Smooth Muscle: An Assessment of Current Knowledge brings us up to date with the use of electron probe analysis and ion selective electrodes, the latter in the laboratory of one of the editors, to verify the unusually high cytoplasmic concentration and activity of chloride in smooth muscle. Perhaps some of the more speculative aspects of this book, such as the postulated binding of calcium to the inner surface of smooth muscle plasma membrane, will eventually also be tested by direct methods.

The book is recommended to those with an interest in smooth muscle and to those wishing to learn about a cell system admirable in its infinite variety.

Andrew P. Somlyo is Professor of Physiology and Pathology and Director of the Pennsylvania Muscle Institute, University of Pennsylvania.

© 1981 Nature Publishing Group